PROCEEDINGS E REPORT
- 45 -

AXMEDIS 2008

Proceedings of the 4th International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution

Workshops, Panels and Industrial Applications

Florence, Italy
17-19 November 2008

Edited by:
Paolo Nesi, Kia Ng, Jaime Delgado

Sponsored by:
DISIT-DSI Lab, University of Florence
Comunicare Digitale
Fondazione Rinascimento Digitale
AxMediaTech s.r.l.

Supported by:
AXMEDIS partners

Firenze University Press
2008

AXMEDIS 2008: proceedings of the 4th International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution. Workshops, Panels and Industrial Applications. Florence, Italy 17-19 November 2008
edited by Paolo Nesi, Kia Ng, Jaime Delgado
ISBN: 978-88-8453-811-6 (online). ISBN 978-88-8453-810-9 (print). © Firenze University Press, 2008

AXMEDIS 2008 : proceedings of the 4th International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution : workshops, Panels and Industrial Applications : Florence, Italy 17 – 19 November 2008 / edited by Paolo Nesi, Kia Ng, Jaime Delgado. – Firenze : Firenze University Press, 2008.
(Proceedings e Report ; 45)
http://digital.casalini.it/9788884538116

ISBN 978-88-8453-811-6 (online)
ISBN 978-88-8453-810-9 (print)

658.514 (ed. 20)

Università degli Studi di Firenze
Firenze University Press
Borgo Albizi, 28, 50122 Firenze, Italy
http://fupress.com

Printed in Italy

AXMEDIS 2008
www.axmedis.org/axmedis2008/
Villa Finaly – Florence, Italy
17-19 November 2008

Sponsored by

Dipartimento di Sistemi e Informatica,
Università degli Studi di Firenze, DISIT Lab.
(Italy)

Associazione Comunicare Digitale (Italy)

Fondazione Rinascimento Digitale (Italy)

AxMediaTech S.R.L. (Italy)

Supported by AXMEDIS partners:

DSI-DISIT, University of Florence (Italy), ACIT - Advance Concepts for Interactive Technology (Germany), AFI - Associazione dei Fonografici Italiani (Italy), ANSC - Accademia Nazionale di Santa Cecilia, Fondazione (Italy), APT - Associazione Produttori Televisivi (Italy), Art In Vision, DigiChannel.net (Italy), AXMEDIATECH Srl (Italy), BBC - British Broadcasting Corporation (UK), DIPITA - Dipartimento di Italianistica, University of Florence (Italy), ELION Enterprises Ltd (Estonia), EPFL - Ecole Polytechnique Federale de Lausanne (Switzerland), ETRI - Electronics and Telecommunication Research Institute (Korea), Eureka s.r.l. (Italy), EUTELSAT S.A. (France), EXITECH srl (Italy), FHGIGD - Fraunhofer Gesellschaft zur Foerderung der Angewandten Forschung E.V. (Germany), HP - Hewlett Packard Italiana S.r.l. (Italy), ILABS - Giunti Labs S.r.l. (Italy), KTU - Kaunas University of Technology (Lithuania), MBI Group (Italy), PKU - Peking University (China), RAI - Radiotelevisione Italiana (Italy), SEJER - representing Bordas and Nathan (France), STRATEGICA (Italy), TEO (Lithuania), TI - Telecom Italia (Italy), Telecom Bretagne (France), TISCALI Italia (Italy), UNIVLEEDS - Interdisciplinary Centre for Scientific Research in Music, University of Leeds (UK), UPC - Universitat Politecnica de Catalunya (Spain), UR - The University of Reading (UK), VRS - Video Reklamos Studija (Lithuania), XIM Limited (UK), FOCUSEEK - PSItrust/focuseek (Italy), SIAE (Italy), Hexaglobe (France), Rigel (Italy), Gesfor (Spain), Pentex (Italy), ALBENIZ Foundation (spain), MAAT-G (Italy), AI2 (Italy), TELSEY (Italy), SDAE - Sociedad Digital de Autores y Editores (Spain)

Table of Contents

AXMEDIS 2008: proceedings of the 4th International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution. Workshops, Panels and Industrial Applications. Florence, Italy 17-19 November 2008
edited by Paolo Nesi, Kia Ng, Jaime Delgado
ISBN: 978-88-8453-811-6 (online). ISBN 978-88-8453-810-9 (print).

Message from the Chairs

Advancements in science and technology have great impacts to everyone. Digital Content is becoming more and more *intelligent* and *cross-media* every day. Media, Multimedia, Cross Media end other definitions are just the evolution of the market while the core substance is the growing complexity of the content and of the completeness of the user experience. For some years now, AXMEDIS has worked in this field in both research as well as large scale industrial applications, with the support of the EC. This is the first AXMEDIS conference after the completion of the project.

The AXMEDIS 2008 International Conference seeks to promote discussion and interaction between researchers, practitioners, developers and users of tools, technology transfer experts, and project managers. The AXMEDIS conference series brings together a variety of participants from the academic, business and industrial worlds, to address different technical and commercial issues. Particular interests include the exchange of concepts, prototypes, research ideas, industrial experiences and other results. The conference focuses on the challenges in the cross-media domain, including production, protection, management, representation, intelligent content, formats, aggregation, workflow, distribution, business and transaction models. Additionally, the conference explores the integration of new forms of content and content management systems and distribution chains, with particular emphasis on the reduction of costs and innovative solutions for complex cross-domain issues and multi-channel distribution.

The AXMEDIS International Conference has been held in the past in Florence (Italy), Leeds (UK), Barcelona (2007). Typically, the conference has 200-250 of attendees from over 20 countries with 50% from research and academic sectors, 40% from the industry, and 10% from government and cultural institutions, etc. The event consists of co-located Workshops, panels, and Tutorials. This year, the conference also hosts additional events of Comunicare Digitale and workshops on cultural heritage contents thanks to Fondazione Rinascimento Digitale and many other colleagues and experts.

This year, the program committee has received an impressive number of submissions for research and applications, industrial panels and workshops. The selection process has not been easy due to the amount of high quality submissions and the limited time slots of the conference. The technical programme produced is very dense with high quality presentations, including a large number of scientific and industrial presentations, industrial panels, workshops and tutorials. This Second volume of the proceedings contains the papers from the workshops, panels, and industrial applications.

We are very grateful to many people without whom this conference would not be possible. Thanks to old and new friends, collaborators, institutions, and organisations who have supported AXMEDIS. A special thank to all the Workshops and Panels organizers. They are really too many to be mentioned in this short note. Thanks also to sponsors and supporters. A very warm thanks to members of the International Program Committee for their invaluable contributions and insightful work even for the industrial papers. Last but not least, many thanks to the many people behind the scene and to all participants of AXMEDIS 2008. We look forward to welcoming you to Florence and wish you an exciting, enjoyable, excellent conference.

General Chair:	Paolo Nesi, DSI-DISIT - University of Florence, Florence, Italy
Programme Co-Chairs:	Kia Ng, ICSRiM - University of Leeds, Leeds, UK
	Jaime Delgado, Universitat Politècnica de Catalunya, Barcelona, Spain

Conference Organising Committee

General chair

Paolo Nesi
DSI-DISIT, University of Florence, Italy
nesi@dsi.unifi.it, www.dsi.unifi.it/~nesi

Programme co-chairs

Kia Ng
ICSRiM – University of Leeds, UK
kia@computer.org, www.kcng.org

Jaime Delgado
DMAG – Dept. Arquitectura de Computadors
Universitat Politècnica de Catalunya, Spain
jaime.delgado@ac.upc.edu, http://research.ac.upc.edu/dmag/

Tutorial co-chairs

Pierfrancesco Bellini and Ivan Bruno
DSI-DISIT, University of Florence, Italy

Local Committee co-chairs

Nicola Mitolo
AxMediaTech, Italy
mitolo@axmediatech.com, http://www.axmediatech.com

International Programme Committee

Web Tv, Iptv, Social Network and Business Tv Experiences. A new era for a crossmedia scenario

Video everywhere! A new scenario is ready to achieve more interest and investment in the multimedia sector. New stars are born: web tv for new contents and platforms; Iptv and the new idea of an evolution of Hd and Vertical Tv; Business Tv to built a community network even for the well known companies itw *(brand Tv)* and Social Network to be always on everywhere.

A concrete meeting to learn more capabilities in a very interesting market.

Chairman
Andrea M. Michelozzi, *Comunicare Digitale*

AXMEDIS 2008: proceedings of the 4th International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution. Workshops, Panels and Industrial Applications. Florence, Italy 17-19 November 2008
edited by Paolo Nesi, Kia Ng, Jaime Delgado
ISBN: 978-88-8453-811-6 (online). ISBN 978-88-8453-810-9 (print).

Projects & Plans for Digital Tv in Europe.

Comunicare Digitale Meeting

Europe are looking for a serious and promptly model for the Dvbt, but is not only a switch over process 'cause involving Dvbh, Hdtv, Iptv and Tv on the Net. How the European government is driving this process? How the Europe consider effective even in South America and United States? Who will guide the process: the Public Television or the Tv Leader in each country? How each country consider this challenge ready to share with the others one? We will describe the situation in Italy, Spain, France and Germany, with a view to Latin America.

The meeting will promote a new scenario even for Comunicare Digitale and the members to approach strongly the next opportunities.

Chairman
Andrea M. Michelozzi, *Comunicare Digitale*

AXMEDIS 2008: proceedings of the 4th International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution. Workshops, Panels and Industrial Applications. Florence, Italy 17-19 November 2008
edited by Paolo Nesi, Kia Ng, Jaime Delgado
ISBN: 978-88-8453-811-6 (online). ISBN 978-88-8453-810-9 (print).

AXMEDIS 2008

Workshops on Industrial Applications

Organised by

DISIT-DSI, University of Florence (Italy)

Coordinated by

Prof. Paolo Nesi, University of Florence (Italy)

Nicola Mitolo, AxMediaTech s.r.l. (Italy)

AXMEDIS 2008: proceedings of the 4th International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution. Workshops, Panels and Industrial Applications. Florence, Italy 17-19 November 2008
edited by Paolo Nesi, Kia Ng, Jaime Delgado
ISBN: 978-88-8453-811-6 (online). ISBN 978-88-8453-810-9 (print). © Firenze University Press, 2008

dContentWare Project:

Combining semantic web and multimedia distribution technologies to realize innovative business models for the provision of digital contents.

Giuseppe Bux
Graphiservice S.r.l.
Bari, Italy
giuseppe_bux@alice.it

Federica Dentamaro
Gius. Laterza & Figli S.p.A
Bari, Italy
dentamaro@laterza.it

Luigi Intonti, Mina Ligorio
Software Design S.r.l.
Bari, Italy
lintonti@softwaredesign.it
mligorio@softwaredesign.it

Ivana Malatesta, Lorella Lamacchia
AI2 – Applicazioni di Ingegneria Informatica S.r.l.
Bari, Italy
malatesta@ai2.it
lorella.lamacchia@gmail.com

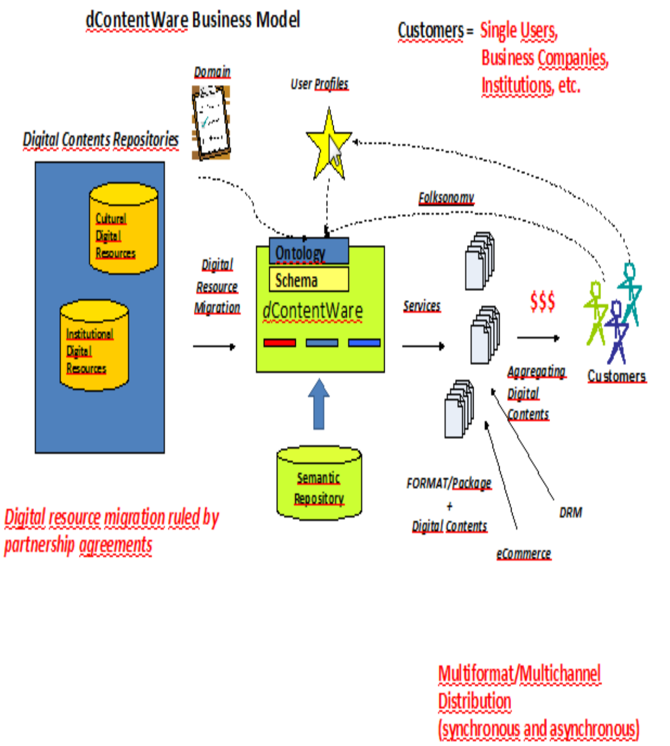
Abstract— **Research on semantic web and multimedia distribution technologies have gained meaningful advances in the recent years; it is now time to experiment, on top of them, innovative business models for the worldwide provision of digital contents: this is the primary goal of the dContentWare project. This paper describes the experience being done in the dContentWare project[1] to reach its primary goal by customizing and combining services provided by JeromeDL, a digital library with semantics and AxMedis, a technological framework for multiformat aggregation and multichannel distribution of digital contents. The proposed business model is being experimented on a concrete use case aiming at deploying a process of semantic migration of bibliographic resources, from the historical archive of Gius. Laterza & Figli publisher into JeromeDL, and their subsequent multiformat aggregation and multichannel distribution through AxMedis.**

Keywords- JeromeDL, Axmedis, digital library, db2rdf, semantic query, social bookmarking, multiformat composition, multichannell delivery, DRM, user profiling.

I. THE DCONTENTWARE BUSINESS MODEL AND THE SET UP OF ITS TECHNOLOGICAL INFRASTRUCTURE

The here proposed dContentWare business model (Figure 1) develops on three deployment frameworks:

1. migration of digital resources from conventional repositories to the dContentWare semantic repository: the migration occurs according to partnership's agreements between the dContentWare consortium and digital contents providers;
2. social bookmarking and semantic searching of digital contents, driven by user profiles;
3. bridging semantic search results towards multiformat and multichannel distribution of digital contents.



[1] The project is co-founded by the government of the Apulia region and is being deployed by an enterprise consortium between the following companies: Gius. Laterza & Figli S.p.a (http://www.laterza.it), Graphiservice Srl.(http://www.graphiservice.it), AI2 Srl., Software Design Srl (http://www.softwaredesign.it)

FIGURE 1. DCONTENTWARE BUSINESS MODEL

The dContentWare infrastructure is built on top of two basic technological platforms:

- JeromeDL[2] : social semantic digital library that makes use of Semantic Web and Social networking technologies to enhance knowledge sharing about digital contents;
- AxMedis[3]: software platform providing services for: authoring multi-format composition models of digital contents; offering configurable research schema of digital contents; aggregating digital contents in predefined combination models; providing multichannel distribution of digital contents by managing Digital Right (DRM) protections.

In the dContentWare business model perspective, JeromeDL acts as the repository of digital contents, while AxMedis acts as the digital contents composer and distributor.

A dContentWare portal (http://www.dcontentware.it), set up in the project, addresses the services provided by JeromeDL and AxMedis, and provides a web service that makes interoperable the JeromeDL repository and the AxMedis composition services.

- A concrete use case of the dContenrtWare business model and its underlying technological framework is currently being deployed and experimented. It is based on the semantic migration of a selected sample of digital resources from the historical archive of Gius. Laterza & Figli S.p.a., one of the leading Italian publishers of humanistic literature, partner of the dContentWare consortium.
- The following sections of this paper describe the customization/implementation process which has been carried out on top of each adopted technological framework, in order to realize and demonstrate the dContentWare business model.

II. DCONTENTWARE: MIGRATING, SOCIALIZING AND SEARCHING DIGITAL RESOURCES IN THE JEROMEDL SEMANTIC REPOSITORY

A. *Migrating*

Digital resource migration is the first process context of the dContentWare business model. It occurs in the frame of a partnership agreement between the owner of an archive of cultural digital resources and the dContentWare consortium.

Usually cultural digital resources are available and managed in conventional repositories, generally in relational databases. Such storage status of digital resources is the first issue to be managed for the concrete succeeding of semantic web technologies. Digital resources need to be extracted from databases available in institutions and made them understandable to semantic web technologies by translating their descriptions in a semantically understandable language, according to the language architecture of the semantic web paradigm [1].

In the dContentWare demonstration use case, the conventional repository of cultural digital resources is constituted by the Historical Archive of Laterza publisher.

The dContentWare migration process takes place in five steps:

1. Analysis and normalization of the logical schema of the original repository: a selection is done of entities and properties involved in the migration process and a resulting entity-relationship model is produced;
2. Specification of a JeromeDL based use case ontology: an ontology model is represented by enriching the JeromeDL ontology model [2] with new properties resulting both from the Analysis step and from emerging design decisions;
3. Design and Building of a transition repository : a normalized transition repository is designed and implemented in a MySql database, by combining the entity-relationship model resulting from the Analysis step, and the new entities/properties emerging from the design of the use case ontology; then a PhP script, with inside MySql queries, extracts resource descriptions from the original archive and moves them into the transition repository;
4. Semantic mapping Db-->JeromeDLOntology: a rdb2rdf mapping specification is produced in D2RQ language [3], in order to map the rdb logical model, of the being migrated archive, to the target JeromeDL rdf based ontology model; the running of the mapping specification, on the D2RQ engine [3], produces a RDF/XML description of the being migrated digital contents;
5. Importing digital contents into JeromeDL: the rdf/xml description of the historical archive is imported directly in the Sesame [4] repository underlying the JeromeDL digital library.

Obviously, JeromeDL provides its own on-line feature to upload resources. In the dContentWare migration process such a feature was used to upload contents components of already migrated resource descriptions, as well to describe and upload single new resources and to edit updating to already migrated/uploaded resources.

[2] The JeromeDL platform (http://www.jeromedl.org) results from a research project managed by Main Library of Gdansk University of Technology [http://www.bg.pg.gda.pl/]and DERI.International [http://www.deri.org/].

[3] The AXMEDIS (http://www.axmedis.org) platform results from a research project co-funded by the European Commission in the frame of IST FP6 and involves about 35 partner as: University of Florence, TISCALI, AFI, SEJER, ILABS, EUTELSAT, HP, Telecom Italia, Telecom Lituania, Telecom Estonia, EPFL, FHGIGD, ACIT, Technical University of Catalonia, University of Leeds, etc.

B. *Socializing*

The original semantic web paradigm [1] was conceived to make machine understandable the web contents. On top of this technological base, a new web is emerging: the collaborative web or web 2.0. Such an evolution is particularly relevant for an innovative business model of digital contents, such as the dContentWare one. It aims at evolving the digital library portals, traditionally conceived as repositories and providers of nude digital contents, towards a new concept of promoters of collaborative generation and sharing of user knowledge about the provided digital contents. Such a new concept mainly characterizes the JeromeDL digital library, so its suitability to support the realization of the dContentWare business model.

The dContentWare users, when acting in the dContentWare/JeromeDL instance, can specify their own private virtual bookshelves (Fig. 2) and their preferred dContentWare digital resources. Each bookshelf is built in terms of a hierarchy of folders of preferred digital resources. Each folder is tagged with concepts extracted from controlled cultural taxonomies originally provided by JeromeDL, as well as from dContentWare own taxonomies resulting from the migration process, as above described. So each folder, with its classification tags, becomes a bookmarking framework of the user preferred digital resources.

Each JeromeDL user is characterized by a FOAFRealm user profile [5], where the friendship relationships of the user with other JeromeDL users are specified. So the user private bookshelf of bookmarked digital contents, can include also the private bookshelves of his friends, so sharing with them their respective preferred digital resources, all tagged by using common taxonomies provided by JeromeDL. This collaborative filtering process is well specified in the SSCF model [6] underlying the bookmarking process in JeromeDL. Other than bookmarking digital resources, the JeromeDL users can comment them and reply to comments of other users in a weblog like feature provided by JeromeDL. An additional JeromeDL feature, on user annotations, concerns their internal representation in RDF, according to the SIOC ontology [7]. It provides a machine understandable representation of user annotations, that can be exported in external more extended weblog environments, such as the WordPress platform, which was adopted by dContentWare in order to animate on-line communities on themes and topics concerned with its provided digital contents.

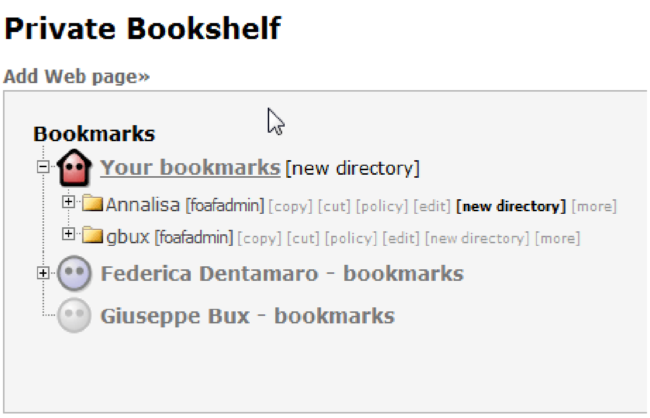


FIGURE 2. USER BOOKSHELF OF BOOKMARKS

C. *Searching*

The semantic web paradigm [1] introduced semantics in resource description. What characterizes and distinguishes JeromeDL from conventional digital libraries is the capability to manage resource searching on a semantic base. Such a searching capability, in JeromeDL, has five foundations:

1. Rdf [9] description of resources: it is based on metadata, whose semantics is unambiguously defined in the JeromeDL ontology [2] and in its ancillary external ontologies; it generates an internal representation of a resource description as an oriented graph;
2. Semantic vocabularies: they include Jonto [10] ontologies of Wordnet vocabulary and worldwide adopted cultural taxonomies, respectively supporting keyword and domain classification of resources;
3. SeRQL: a powerful RDF query language, provided by Sesame [3], enabling the specification of fine grain criteria in semantic queries, so allowing to rapidly reach resources matching the user search objectives;
4. Regular Expressions: foundation of computer language theories; they are adopted in JeromeDL to specify patterns of queries expressed in Natural Language (NL);
5. Templates of queries: predefined templates of SeRQL and NL queries coded in the system; they allow an user to select the NL query more concerned with his search objective and to instantiate it with search criteria. On the base of the selected and instantiated NL query template, the system in turn selects, instantiates and makes running, on the Sesame [4] repository, the proper SeRQL query template, so extracting the resources matching the search criteria specified in the NL query template.

On the above foundations, it is possible to specify, in the system, effective and user friendly query templates, which enable users to dynamically compose and launch very powerful semantic queries acting on semantic metadata of the whole oriented graph that characterizes a RDF repository.

Obviously, as in the conventional digital libraries, JeromeDL provides string based search mechanisms, based on full text indexing engines, but, as commonly known, the limitation of such an approach is the large grain of the generated search results, because of lack of semantic meaning of research criteria.

The semantic query capability of JeromeDL is particularly suitable to realize the dContentWare Business Model. It in facts allows the dContentWare query designer to customize JeromeDL with templates of rdf semantic queries, based on the resource aggregation criteria of the AxMedis composition services.

III. dCONTENTWARE: MULTIFORMAT COMPOSITION AND MULTICHANNEL DISTRIBUTION OF DIGITAL CONTENTS

A. Preliminary remarks

A critical issue of the dContentWare Business Model is the identification of appropriate categories of services, providing added value to the final users.

In order to define and experiment presentation forms of such added value services, a fundamental aspect is the capability to aggregate digital resources on semantic base and to compose them in predefined presentation formats, as well as the capability to distribute them on different distribution channels, by assuring DRM protections.

In the specific framework of the dContentWare Business Model, an added value service is characterized by:

- contents: i.e. the digital resources the service works on;
- distribution: i.e. format, channel, scheduling implemented by the service;
- administration : i.e management of fee, rights,... of digital resources.

Taking into account these preliminary remarks, the choice of Axmedis platform has turned out to be extremely functional for the purpose of the dContentWare Business Model. It allows to test the different functionalities of an added value service, according to its envisaged concept in the dContentWare Business Model.

B. Working in dContentWare/AxMedis

The dContentWare/Axmedis users , entering in the system (Fig. 3), can subscribe a specific service by selecting it in the list of the current active ones. Moreover the user can define specific settings related to selection and fruition of digital contents. Each dContentWare/Axmedis user describes in the system his own profile, which is functional to the system for the proper presentation of services to the user itself.

In order to test the features of the different services, being configured for the dContentWare purpose, and thei related user subscriptions, all information generated in the services are stored in a MySQL Database, that is accessible from the Axmedis Content Processing component (AXCP).

The service setting process includes parameters such as:

- scheduling: daily, weekly, etc.
- distribution channel: email, podcast, etc.
- reference domain/category: literature, history, ...
- subject: ancient history, modern history, ...
- source: Laterza publisher archive, Jeromedl repository,
- license: “pay to play,” subscription, ...
- admitted user rights: Modify, copy, ...

In the dContentWare Business Model perspective, AxMedis does not manage its own repository of digital contents, but it gets resources from the JeromeDL semantic repository.

According to the predefined schedule of the service , the AXCP component executes a procedure for its deployment.

The first step of the AXCP procedure manages the acquisition of the digital resources from JeromeDL, via the “Transducer” web service, provided by the dContentWare portal (see next section).

A second step of the procedure, running as from its predefined schedule, composes the resources, get from JeromeDL, into new Axmedis objects. The composition is ruled by predefined logics and formats, specified for the service by filled in SMIL (Synchronized Multimedia Integration Language) [10] templates or by customized html pages. The filled in SMIL templates constitute the requirements scenarios for semantic searching of digital contents in the JeromeDL repository. Dublin Core [11] metadata are adopted to describe the AxMedis object description.

Moreover, based on the user subscriptions to each service, the AXCP component creates a DRM license for each Axmedis object, where digital rights are specified, related to the user, the composer and the adopted distribution device (both on the web and on mobile systems).

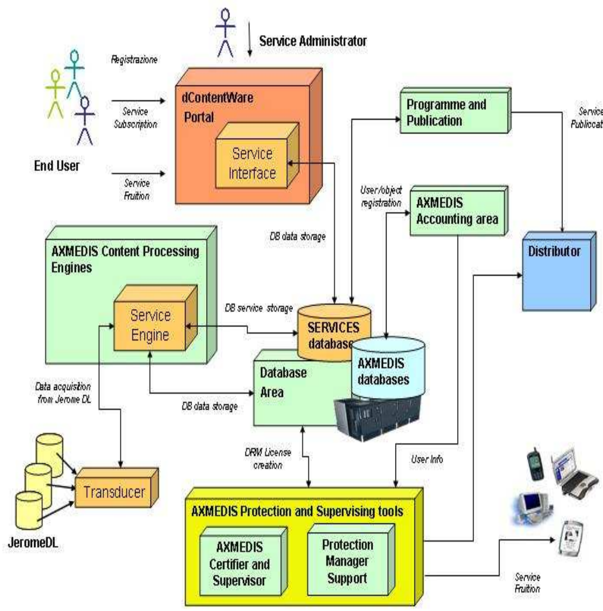


FIGURE 3. INTERACTION BETWEEN AXMEDIS AND dCONTENTWARE

C. *The use case of dContentWare/AxMedis services*

The current use case of dContentWare/AxMedis services is being deployed on bibliographic resources migrated from the Historical Archive of Laterza publisher into the dContentWare/JeromeDL semantic repository (see previous section).

The AxMedis service configuration has been defined according to the rules defined in the partnership agreement between the dContentWare consortium and the owner/provider of the digital contents repository.

Such rules concern the kinds of provided digital resources and the related distribution licenses. The established kind of digital resources is driving the specification of resource aggregation criteria and composition formats, and in turn semantic search criteria in the JeromeDL repository.

Examples of semantic search/aggregation criteria are:

- resources written/published/edited by a given author/publisher/editor in a given temporal period;
- resources within a being required cultural domain and with a specific key subject;
- resources whose contents refers to a cultural domain and related to a given historical period;
- resources bookmarked by dContentWare/JeromeDL users within a given cultural domain and/or with a given key subject;
- resources belonging to a given resource collection;
- etc..

The being initiated life cycle of the dContentWare/Axmedis services will suggest additional retrieval/aggregation criteria.

Such retrieval/aggregation criteria drive the specification of templates of semantic queries in JeromeDL. These are instantiated at deployment time of an Axmedis service, by enacting the "Transducer" web service provided by the dContentWare portal (see next section).

IV. DCONTENTWARE: BRIDGING DIGITAL CONTENTS FROM JEROMEDL SEMANTIC REPOSITORY TO AXMEDIS MULTIFORMAT COMPOSITION SERVICES.

As until now described, the dContentWare Business Model is supported by JeromeDL and AxMedis, respectively a digital library and a resource composition and distribution framework. In order to assure interoperability between such environments, two additional technological components were developed in the project:

- the dContentWare Portal, which allows the interactive access to the whole dContentWare environment;
- a "Transducer" of JeromeDL resources versus AxMedis handling services.

A. *The dContentWare Portal*

The dContentWare portal (www.dcontentware.it) is the main front-end for the user. It introduces the dContentWare Business Model concepts and provides direct access to the dContentWare services, available through JeromeDL and AxMedis, as well as trough the dContentWare ancillary Wiki and Weblog environments.

The JeromeDL services are exposed within an own Jeromedl portal, whose link (www.dcontentware.it/jeromedl) is provided to the user by the dContentWare portal.

The Axmedis services, differently from JeromeDL, are not exposed within an Axmedis portal. The underlying functionalities are presented in an user "Subscription" page provided by the dContentWare portal. Here all the AxMedis composition and distribution procedures of digital contents are listed.

The user can define some settings for each procedure, and then he can subscribe the related AxMedis services, offered him according to specific DRM licenses.

The settings of an AxMedis procedure include different inputs:

- user settings: they are provided at the subscription stage of the procedure and include information such as profile user descriptions, delivery time of the services results, etc.
- AxMedis editor settings, regarding contents aggregation models.

On the base of such settings, the system launches semantic queries in JeromeDL, by instantiating therelated RDF query templates.

A web service application manages the communication between the dContentWare portal and AxMedis, by exhanging data coded in XML Such XML datainclude the list of the available AxMedis procedures and the list of the procedures subscribed by a specific user.

The portal, moreover, provides a web page that, on the base of parameters provided by AxMedis, is able to dynamically produce forms that allows the setting of contents aggregation and distribution parameters. These forms are returned back to AxMedis by an additional web service.

B. *The JeromeDL-AxMedis Transducer*

Digital contents being composed and distributed through AxMedis are available and retrieved in the JeromeDL digital library. JeromeDL supports different kind of queries (namely: simple, advanced, semantic), where the more suitable for the AxMedis purpose are the semantic queries. JeromeDL query results are returned in RDF [8] format, representation that is not supported by AxMedis. The current 2.1 release of JeromeDL, adopted in dContentWare, provides its query framework via user interface, as well as it provides mechanisms for distributed queries on a network of JeromeDL instances, by using the HyperCup [10] communication protocol. AxMedis in turn provides a technological framework to enact its services via web services applications. In order to overcome the different technological protocols of JeromeDL and Axmedis, a web service was implemented to make them interoperable. This is the "Transducer" web services (Figure 4): it is interfaced with AxMedis and it allows to launch queries on JeromeDL and to give back results at Axmedis, without using the HyperCup communication protocol adopted in JeromeDL.

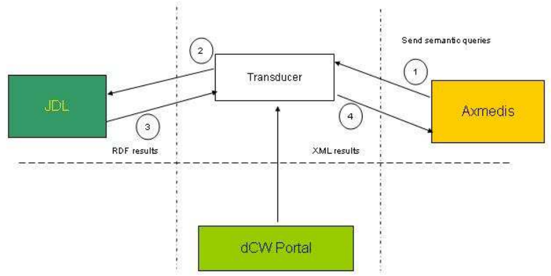


Figure 4. dContentWare Transducer web service

The "Transducer" web service then, by launching queries in JeromeDL and by capturing the query results, in programmatic way, provides the composition and fruition procedures of AxMedis with the required resources.

Relevant feature of the "Transducer" web service is the adoption of the same JeromeDL mechanism to launch a query, as well as the making understandable, to AxMedis, the query results from JeromeDL. The web service methods, in facts, launch a query by generating the related JeromeDL URI, as it would be generated in the JeromeDL user interface context. JeromeDL gives back query results by providing features for representing them in RDF compatible code (XML, n3, ntriple, etc.). The "Transducer" enacts the feature to represent RDF in XML code: this provides a verbose description of the JeromeDL history about a resource. The "Transducer" intercepts the RDF/XML stream, then it captures the pure information about a resource, such as "title", creator", abstract, and URIs of its contents components, and it composes it in a simplified XML document submitted to AxMedis.

Moreover, it is possible to properly configure the "Transducer" in order to allow it to include, in the simplified XML description of a JeromeDL resource, additional rdf tags, without modifying the "Transducer" code.

References

[1] Tim Berners-Lee, James Hendler and Ora Lassila: The Semantic Web - http://www.sciam.com/article.cfm?id=the-semantic-web

[2] Sebastian Ryszard Kruk: JeromeDL Ontology 1.0 Specification - http://www.jeromedl.org/structure/spec/index.html

[3] Chris Bizer: The D2RQ Plattform - Treating Non-RDF Databases as Virtual RDF Graphs - http://www4.wiwiss.fu-berlin.de/bizer/d2rq/

[4] Open RDF.org: Sesame RDF Schema Querying and Storage - http://www.openrdf.org/

[5] Slawomir Grzonkowski , Sebastian Kruk : D-FOAF: Role-based Access Control Standard - http://www.foafrealm.org/documentation/AccessControl/

[6] Adam Gzella: Social Semantic Collaborative Filtering - http://library.deri.ie/resource/FUxGypZk

[7] Uldis Bojars, Jhon G. Breslin: SIOC Core Ontology Specification - http://rdfs.org/sioc/spec/

[8] W3C: Resource Description Framework (RDF) - http://www.w3.org/RDF/

[9] Jonto project: Java bindings for ontologies - http://sourceforge.net/projects/jonto

[10] SMIL: Syncronized Multimedia Integration Language - http://www.w3schools.com/smil/default.asp

[11] Dublin Core: Dublin Core Metadata Initiative - http://dublincore.org/

[12] Slawomir Grzonkowski , Sebastian Kruk : HyperCup Lightweight Implementation - http://library.deri.ie/resource/AlZpcTQx.

Interactive Contents Protection in Digital Terrestrial Television

M. Baldi [1], G. Cancellieri [2], E. Gambi [3], G. Rascioni [3], S. Spinsante [1]

[1] ArieLAB Srl
Ancona, Italy
{m.baldi, s.spinsante }@arielab.com

[2] Centro Radioelettrico Sperimentale G. Marconi
Ancona, Italy
giovanni.cancellieri@centromarconi.com

[3] Università Politecnica delle Marche
Ancona, Italy
{e.gambi. g.rascioni}@univpm.it

Abstract—**Among the most important innovations brought by the introduction of Digital Terrestrial Television, interactive applications have a prominent role. The DVB MHP standard adopts Java applications, named Xlets, for implementing interactive services on the TV platform. When conveyed over a broadcast channel, the Java bytecode is exposed to the risk of decompilation, that may result in infringement of copyright and intellectual property. In this paper, we study the adoption of Java code obfuscation techniques in DVB MHP applications in order to prevent attackers from recovering a valid source code for interactive contents. We also assess the performance of obfuscated bytecode through its execution on a commercial digital terrestrial television receiver.**

I. Introduction

The introduction of Digital Terrestrial Television (DTT), and its fast spreading, has opened the way to a new possibility in the delivery of digital Audio/Video (A/V) information and data. As a matter of fact, the MPEG-2 compliant Transport Stream (TS) adopted in Digital Video Broadcasting – Terrestrial (DVB-T) technology [1] allows to carry not only properly encoded and compressed A/V Elementary Streams (ESs), but also additional metadata. They may consist in different resource files (text, images, and so on) and in interactive applications that are based on the Java language, in the specific case of interactive DVB-T transmissions.

Given the intrinsic nature of digital content, receivers can easily copy, modify, and redistribute it, if a proper protection system is not applied. Traditional Conditional Access (CA) systems can protect broadcast data during transmission, but protection is completely removed once the data are descrambled at the receiver. Provided that Set Top Boxes (STBs) equipped with storage capacities are spreading on the market, the digital content may be copied and redistributed, also with the aim of a non personal use, once it has been legally descrambled, and stored on a hard disk.

Besides ensuring protection of the digital content, it may be required to set different rights on different data: this can be accomplished through a Digital Rights Management (DRM) system. Several DRM solutions exist and have been proposed [2]-[4] in a broad range of different application contexts; some have been specifically developed for the digital television [5], [6]. In [7], the authors discuss a robust DRM system implemented through the DVB Multimedia Home Platform (MHP) [8], a smart card, and a fast encryption algorithm. However, the proposed scheme requires a suitably configured smart card, which may be not available in practice, at least on a large scale.

In the early 2008, the Digital Video Broadcasting Consortium has released the DVB CPCM specification [9], which defines a system for Content Protection and Copy Management of commercial digital content delivered to consumer products and home networks. CPCM is conceived to manage content usage, from acquisition into the CPCM system until final consumption, or export from the CPCM system, in accordance with the particular usage rules of that content. Although the functionality targeted for DVB CPCM is much less ambitious than that of a full DRM system, the scope envisaged by the DVB consortium is for end-to-end protection of commercial digital content in all processes, from the point of acquisition by the consumer through to the point of consumption. Among the possible sources for commercial digital content, broadcast (e.g., cable, satellite, and terrestrial) and Internet-based services, packaged media, and mobile services may be included. CPCM is intended for use in protecting all types of content - audio, video and associated applications and data. DVB CPCM can interface a DRM system, by translating the DRM system license and user rights into a specific set of so called CPCM objects. By this way, the DVB Consortium did not recommend the choice of a specific DRM system for Digital Terrestrial Television, leaving the operators the possibility of performing different choices, based on specific needs or requirements.

Actually, it is well known that the strength and complexity of any security solution should be traded off with the true sensitiveness of data and information that is necessary to protect. Even if many robust and hard-to-brake solutions may be conceived, often the final choice is performed on the basis of the minimum computational requirements needed. As a consequence, a common scenario in DVB-T provides scrambling of the A/V material through a CA system, and obfuscation of the interactive applications carried by the TS by means of software, ad-hoc, tools. Code obfuscation is often preferred for protecting interactive contents since it is able to ensure a good level of security against copyright infringement without yielding increased complexity of the receiver.

In this paper, we study the effects of Java code obfuscation techniques when applied to MHP interactive applications, that represent a particular class of Java applications. We also provide some performance evaluations about the execution of obfuscated applications by commercial DTT STBs. The paper is organized as follows: in Section II we introduce the principles of data broadcast in digital terrestrial television; Section III is devoted to MHP interactive applications and their security; in Section IV the most common code obfuscation techniques are reviewed; Section V reports the results of experimental tests, and Section VI concludes the paper.

II. Digital Terrestrial Television

According with the DVB-T standard [1], the audio and video contents for the broadcast transmission are encoded in MPEG2 elementary streams and encapsulated in a DVB compliant TS [10]. Following the ETSI EN 301 192 standard [11], a predefined set of description tables must be included in the transport stream for creating a digital television channel with all the required indexes.

For this purpose, the Network Information Table (NIT), the Service Description Table (SDT), the Program Association Table (PAT) and the Program Map Table (PMT) are cyclically inserted in the transport stream, and form the DVB Service Information (SI) content that links together all the elementary streams within the same transport stream.

Digital terrestrial television also provides interactive services, and interactivity is implemented by means of Java applications compliant with the DVB MHP standard [8]. For this reason, a suitable mechanism has been implemented for indexing and broadcasting Java interactive contents within a DVB transport stream.

As a matter of fact, the DVB-MHP standard provides a further table, named AIT (Application Information Table), containing full information on the data broadcast and the initial state of applications carried by a given TS.

In addition, the set of classes forming each interactive application is cyclically transmitted by means of an object carousel generator (or MHP playout system) able to encapsulate software packets in a Digital Storage Media - Control and Command (DSM-CC) file system [12]. Such file system is split into modules, then into sections and, finally, into transport stream packets, that are multiplexed with audio, video and other data contents.

The PMT of an interactive digital television service contains the reference to the AIT and DSM-CC streams corresponding to the same service, so the user terminal is able to associate interactive applications to television services.

The data broadcast is implemented in a “carousel” fashion, that is, by cyclically transmitting the whole set of Java classes, in such a way that the user terminal is able to fetch the entire file system content independently from the instant at which it is turned on.

This, together with the broadcast nature of the radio channel, gives easy access to interactive applications on air, and exposes them to attempts of reverse engineering. In fact, suitable tools allow to decode DSM-CC contents and to rebuild the whole file system associated to a digital television service [13]. Thus, by means of a common digital terrestrial receiver, an eavesdropper could recover the whole set of Java classes forming an interactive application. Once available, the Java bytecode could be decompiled through suitable and easy to find tools, thus obtaining a valid source code for the interactive application, and allowing theft of intellectual property and copyright infringement.

III. The Multimedia Home Platform

A. *Java and MHP*

Interactive applications, maybe one of the most important and innovative features of the DVB-T technology, are called Xlets and are written in a subset of the Java language, the Multimedia Home Platform [8]. It is an open standard, resulting from an ETSI initiative, that allows the development of interactive applications for the Digital Terrestrial Television, in a way independent from STB hardware constraints.

The MHP software stack is a complex and modular architecture: standardized Application Program Interfaces (APIs) are used internally, to exploit the dependencies among the various components. Besides normal Java APIs, many APIs for TV applications are included in the MHP specification. The Java APIs include graphics components, a User Interface events component, an inter Xlet communication API (typically built on a proprietary API), and the Return Channel component. Scarce resources, like a return channel interface implemented through a Public Switched Telephone Network (PSTN) modem, are handled by means of the DAVIC resource notification API. The Conditional Access (CA) component, used to access and decode scrambled elementary A/V streams, is configured to directly interface the hardware MPEG decoder, mainly because of efficiency issues. A basic component is the resource manager, that provides the other components a framework for sharing scarce resources, and is exposed through the DAVIC resource notification API.

The Interactive Broadcast and Internet Access profiles of MHP include support for return channel, Web browsing and e-mail client. Thanks to these features, the use of a return channel interface in MHP is similar to the use of an IP connection in a classic Java application. The MHP 1.0.*x* specification requires support for HTTP 1.0 and Domain Name System (DNS) over the return channel, on top of the basic Transport Control Protocol (TCP) and User Datagram Protocol (UDP), but any other component is optional. MHP 1.1 adds support for HTTPS, i.e. the secure version of HTTP, but other protocols, such as SMTP or File Transfer Protocol (FTP) are not mandatory. MHP 1.1, originally published in June 2001, introduced smart card reader APIs, to support a low level communication with smart cards, typically used by CA systems; MHP 1.1.2 adds further features, to use smart cards for encryption and user authentication. The MHP 1.1 smart card reader API was based on the Open Card Framework (OCF); in MHP 1.1.2 it is replaced by Security And Trust Services API (SATSA) [14], already supported by Java for mobile terminals.

B. Security in MHP

The topic of security in MHP may get a number of different meanings [15]. First of all, security may be intended as reliability, i.e. to ensure that applications do not cause problems for the STB middleware, when executing. Then, security may be intended by broadcasters as preventing unauthorized people from getting access to content they have not paid for. This issue is typically solved by resorting to CA systems, encryption, and scrambling techniques.

Further, in MHP, there is also the problem of authenticating downloaded applications, in such a way as to be sure they can perform only the operations they are allowed to do. Xlets may be digitally signed by network operators, in order to allow the receiver to check that they have not been modified. By such means, signed applications are granted any additional permission of using the STB resources, besides those basic resources granted by default to any, even unsigned, application.

Signing of MHP applications is based on X.509 certificate chains [16]. The process is built upon three main steps:

- integrity verification, by using the hash values as checksums for files and directories composing the application;
- signing of application, by verifying that the hash values were calculated by the network operator or the application provider;
- signature verification at the receiver, through the X.509 certificate received.

Different sets of files are used at each step, and they must be included in the Object Carousel carrying the application itself. Just creating a hash file is not enough, though. The broadcaster signs hash files using a digital signature algorithm, the public key of which is transmitted to the receiver in a X.509 certificate, carried inside the file system of the Object Carousel.

Practical implementation of hash files in the MHP context may be complicated, because of the limited CPU power available at the receiver, and because of the latency due to the loading of hash files from the Object Carousel. Given a digest value computed for all the objects (files or directories) in a group, all the objects in that group must be loaded from the Carousel in order to check the hash value, and this may determine unacceptable latencies.

Moreover, the security framework based on X.509 certificates and hash files is aimed only at guaranteeing integrity and authentication of data broadcast contents. When protection of the transmitted data must also be ensured against copyright infringement, alternative techniques can be adopted, such as code obfuscation by means of suitable software tools.

IV. JAVA CODE OBFUSCATION

The Java programming language is an object oriented programming language that descends from the C++ language and is aimed at guaranteeing interoperability of the same software packages over different hardware architectures.

For this purpose, applications written in Java are not compiled into the native binary code of each architecture, but instead rely on a suitable middleware that implements hardware abstraction, and that is called the Java Virtual Machine (JVM).

The Java source code is compiled into an intermediate language, called *bytecode*, that is then executed by the JVM. The Java bytecode is formed by a set of simple instructions that represent an abstraction with respect to native instructions of each architecture, and depends on a predefined set of libraries. Each JVM must expose the same set of classes and methods, in such a way that bytecode compatibility is preserved among different platforms.

On the one hand, the Java bytecode is a powerful mean to ensure software interoperability and to implement the Java concept of “compile once run anywhere”. On the other hand, however, the Java bytecode preserves most of the structure of the original source code of an application, so it allows to rebuild it from its compiled version. Several Java decompilers are now available, and their success rate is significantly higher than that of similar tools for traditional programming languages (as C and C++).

This fact, jointly with the ease of fetching the compiled versions of DTT interactive applications, exposes software houses working on them to the risk of copyright infringement and source code piracy.

In order to avoid (or, at least, reduce) such risk, suitable techniques can be adopted that aim at producing an *obfuscated* Java bytecode, that is less prone to be decompiled into source code [17]-[19].

Obfuscator tools aim at making difficult, or even impossible, reverse engineering on the Java bytecode. The most common practices adopted for this purpose are: i) altering the source code control flow by inserting overabundant branches and variables substitutions; ii) renaming classes and class members with names that are meaningless to humans, and iii) removing source file names, comments and debug instructions.

Moreover, several obfuscation techniques – such as renaming to shorter names – help to reduce the size of compiled classes, and may result, as a side effect, in code optimization. This may produce a performance improvement when Java applications are executed. The goals of obfuscation and optimization are different, but the techniques used to accomplish such goals are similar, so it makes sense for the same module or tool to serve both purposes.

However, as it will be evident by the test results reported in the next section, code optimization introduced by obfuscators may not produce any improvement when applied on very small applications, as MHP Xlets. In this case, the introduction of branches and variable substitutions may even produce worse performance.

At present, both commercial and open source Java obfuscation tools are available. Obfuscation does not require access to the source code (it can be directly executed on the Java class files) and, depending on the tool adopted, it can be done by means of a suitable Graphical User Interface (GUI), through a configuration file or via the command line.

V. Experimental Tests and Results

In order to verify the effectiveness of interactive contents protection via Java code obfuscation, we have considered a standard MHP application, that is the freely available "MHP Tester" application [20]. The MHP Tester application represents a useful benchmark since it is able to test some common features of MHP compliant digital television receivers.

The MHP Tester has been compiled into bytecode, and then two obfuscated versions of it have been produced by means of the trial versions of two commercial code obfuscators: DashOPro 4.1 [21] and Klassmaster 5.1 [22]. The reason of our choice relies in the fact that these obfuscators support the compilation of Xlets, while other tools need a more involved configuration for complying with digital terrestrial television standards. A screenshot of the MHP Tester application is reported in Fig. 1.

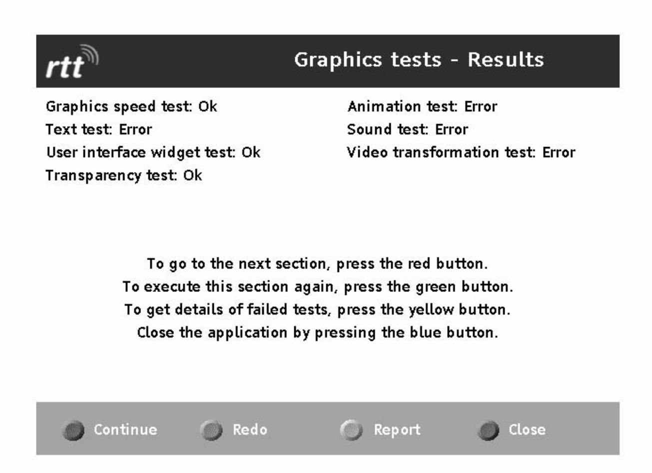


Fig. 1. The MHP Tester application.

This application runs a predefined set of tests on the digital terrestrial television receiver and measures, among others, the following performance indicators: i) number of rectangles drawn in a second, ii) number of ovals drawn in a second, iii) number of lines drawn in a second, iv) number of text strings drawn in a second, and v) frame per second in rendering animations.

Each test produces a video output and the corresponding performance measure. An example of test execution is shown in Fig. 2.

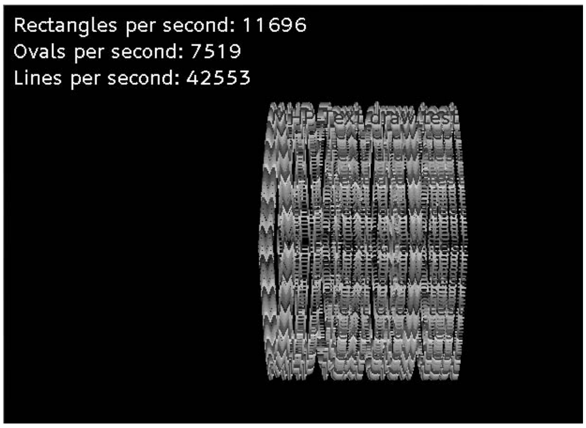


Fig. 2. Example of MHP test.

We have verified that both the obfuscators considered are able to produce a bytecode that prevents the action of decompilers.

Fig. 3 shows a sketch of the original source code, that is characterized by high legibility and presence of comments.

```
/**
 * Constructs the Xlet
 */
public MHPTesterXlet()
{
    Util.setDebugOutput(true);
    Util.debug("MHPTesterXlet constructor");
    context = null;
    currentState = STATE_LOADED;
    runTestModule = null;
    Util.debug("MHPTesterXlet constructor
done");
}
```

Fig. 3. Sketch of the original source code.

When the bytecode is obfuscated by means of DashOPro, and then decompiled into source code, the result assumes the form shown in Fig. 4: the method names are preserved, as well as strings, but variable names have been replaced with meaningless ones.

```
public MHPTesterXlet()
{
    Util.setDebugOutput(true);
    Util.debug("MHPTesterXlet constructor");
    a = null;
    b = 0;
    d = null;
    Util.debug("MHPTesterXlet constructor
done");
}
```

Fig. 4. Decompiled source code from DashOPro obfuscated

bytecode.

Klassmaster is able to reach a higher level of obfuscation: even method names and strings are scrambled in the decompiled source code, shown in Fig. 5.

```
public a()
{
    s.a(true);
    s.a(z[18]);
    a = null;
    b = 0;
    h = null;
    s.a(z[17]);
}

[...]
as[17] =
"U?\037`xk\003*FEt\022;\024~w\031<@om\024;[o8
\023 Zx";
z = as;
```

Fig. 5. Decompiled source code from Klassmaster obfuscated bytecode.

In both cases, the decompiler has not been able to rebuild the correct tree of classes that form the application, and re-compilation into a valid executable has been prevented.

As a last verification, we have assessed the performance of the obfuscated bytecode in terms of the parameters measured by the MHP Tester application.

We have run three versions of the MHP Tester Xlet on an MHP compliant digital terrestrial television receiver [23]: the first application has been compiled without obfuscation, while for the other two we have produced obfuscated bytecode through the considered tools. The results obtained have been averaged by repeating 10 times each experiment, and the corresponding values are reported in Table I.

TABLE I
PERFORMANCE ASSESSMENT WITH AND WITHOUT OBFUSCATION

Measured Parameter	Without Obfuscation	DashOPro	Klassmaster
Rectangles per second	354.5	342.1	350.8
Ovals per second	217.5	217.5	218.5
Lines per second	5836.9	5651.4	5575.6
Text strings per second	201.2	193.5	199
Animation FPS	12.2	12	12

Performance figures do not vary in a substantial way when considering obfuscated and not obfuscated Xlets. However, we can observe that, in the case of very small Java applications (as Xlets are), the code optimization which should result from obfuscation is not effective: the application compiled without obfuscation gave the best results in terms of execution performance.

As concerns the considered obfuscators, we notice that, under the performance viewpoint, DashOPro is able to produce a better optimized bytecode with respect to Klassmaster. This is obtained at the cost of a less aggressive obfuscation that, however, is still able to prevent decompilation.

VI. CONCLUSION

The broadcast nature of digital terrestrial television transmissions makes their interactive content exposed to copyright infringement through bytecode decompilation, and Java code obfuscation is often used in order to prevent it.

With respect to more involved techniques, obfuscation has the advantage of avoiding increased complexity both at the transmitter and at the receiver side.

By testing different commercially available solutions, we have verified that Java code obfuscation may adequately provide protection to DVB MHP applications, since it prevents the usage of decompilers. This feature, joint the low complexity associated to obfuscators, makes them suited to the peculiar MHP context. However, obfuscation may also yield performance degradation when obfuscated applications are executed on a digital terrestrial television receiver. In any case, the performance loss remains limited, so code obfuscation is confirmed as a promising technique for protecting interactive contents from the risk of theft of their intellectual property.

As a further development of this activity, we are currently investigating the possibility of applying open source solutions, in order to provide a common framework for Java code protection in the MHP scenario, where the limits of computational and processing resources at the receiver must be taken into account.

REFERENCES

[1] "Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television," ETSI EN Std. 300 744 (v. 1.5.1), Nov. 2004.

[2] Automating Production of Cross Media Content for Multichannel Distribution (Axmedis), see www.axmedis.org.

[3] W. Zeng, H. Yu, C.-Y. Lin, *Multimedia Security Technologies for Digital Rights Management*. Elsevier, 2006.

[4] Windows Media Digital Rights Management (see http://www.microsoft.com/windows/windowsmedia/forpros/drm/default.mspx).

[5] TV-Anytime, available at http://www.tv-anytime.org.

[6] MPEG-21 Part 4, Intellectual Property Management and Protection (IPMP) available at http://www.chiariglione.org.

[7] C.-K. Liang, C.-C. Liu, H. H. Chen, "A Robust DRM System on the DVB Multimedia Home Platform," in *Proc. IEEE Cons. Comm. and Networking Conf. 2006*, Las Vegas, Nevada, USA, pp. 605 – 609.

[8] "Digital Video Broadcasting (DVB) – Multimedia Home Platform," ETSI TS Std. 102 812 (v. 1.1.1), June 2003.

[9] Digital Video Broadcasting Consortium, "Content Protection & Copy Management Specification; Introduction," DVB Document A094 Rev. 2 Part 0, February 2008.

[10] "Information technology - Generic coding of moving pictures and associated audio information - Part 1: System", International standard ISO/IEC 13818-1, 1996.

[11] "Digital Video Broadcasting (DVB); DVB specification for data broadcasting," ETSI EN Std. 301 192 (v. 1.4.2), Apr. 2008.

[12] "ISO 13818-6 Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - Part 6: Extensions for DSM-CC", International standard ISO/IEC 13818-6, 1998.

[13] libdsmcc, available at http://sourceforge.net/projects/libdsmcc/

[14] Security And Trust Services API for J2ME - SATSA 1.0, Java Specification Request JSR 177, July 2004 (available at: http://java.sun.com/products/satsa).
[15] Morris S., Smith-Chaigneau A., ``Interactive TV Standards: a guide to MHP, OCAP and JavaTV," Chap.13, Focal Press, 2005.
[16] R. Housley, W. Ford, W. Polk, D. Solo, Internet Engineering Task Force, “Internet X.509 Public Key Infrastructure Certificate and CRL Profile,” Request For Comments RFC 2459, January 1999.
[17] S. W. Shah, P. Nixon, R. I. Ferguson, S. R. ul Hassnain, M. Naeem Arbab, and L. Khan, “Securing Java-Based Mobile Agents through Byte Code Obfuscation Techniques,” in *Proc. IEEE INMIC '06*, Islamabad, Pakistan, Dec. 2006, pp. 305 – 308.
[18] J.-T. Chan and W. Yang, “Advanced obfuscation techniques for Java bytecode,” Journal of Systems and Software, vol. 71, pp. 1 – 10, Apr. 2004.
[19] J. M. Memon, S. ul Arfeen, A. Mughal and F. Memon, “Preventing Reverse Engineering Threat in Java Using Byte Code Obfuscation Techniques,” in *Proc. IEEE ICET 2006*, Peshawar, Pakistan, Nov. 2006, pp. 689 – 694.
[20] MHPTester, available at http://sourceforge.net/projects/mhptester.
[21] DashO Java Obfuscator, available at http://www.preemptive.com/products/dasho/index.html.
[22] Zelix KlassMaster, available at http://www.zelix.com/klassmaster/download1.html.
[23] Advanced Digital Broadcast (ADB), i-CAN 2000T DVB-T MHP compliant receiver (http://www.adbglobal.com/)

A Framework for Digital Watermarking Next Generation Media Broadcasts

Dominik Birk†* Seán Gaines* Christoph Wegener†

Abstract—**The Internet presents a problem for the protection of intellectual property. Those who create content must be adequately compensated for the use of their works. Rights agencies who monitor the use of these works exist in many juristictions. In the traditional broadcast environment this monitoring is a difficult task. With Internet Protocol Television (IPTV) and Next Generation Networks (NGN) this situation is further complicated.**
In this work we focus on Digitally Watermarking next generation media broadcasts. We present a framework which provides the ability to monitor media broadcasts that also utilises a Public Key Infrastructure (PKI) and Digital Certificates. Furthermore, the concept of an independent monitoring agency, that would operate the framework and act as an arbiter, is introduced. We evaluate appropriate short signature schemes, suitable Watermarking algorithms and Watermark robustness. Finally, the application of the proposed framework in other related scenarios is discussed.

Keywords: Next Generation Networks, broadcast monitoring, public key watermarking, IPTV, PKI, short signature

1 Introduction

The global acceptance of revolutionary services, such as IPTV [1] and NGN [2], is in the state of growth and impels broadcasters and media creators to evolve their existing infrastructures and services. IPTV, for instance, shows the potential to bring interactive content to the masses. Consumer applications such as interactive TV (iTV) [1] changes the modality of the TV set from being the source of one-way passive entertainment to a two-way interactive entertainment and communications model.
With the transition to digital media streams received over the Internet, new challenges loom. Today, the practices of recording, distribution and copying multimedia content is easy and straightforward [3]. Due to these facts, it is more and more difficult to enforce copyright and to safeguard intellectual property for broadcast media.
Digital Watermarking [4], which may be considered a form of steganography [5], attempts to address this problem by embedding information within the digital signal. The primary use of Watermarking is on digital signals that encode audio, image or video content. However, Digital Watermarking may be used for a wide range of applications such as copyright protection [6], fingerprinting [7], broadcast monitoring, advertising monitoring [8] and communication over covert channels [9]. It is debatable whether traditional Watermarking systems, which are based on disclosure of the key needed to embed and to detect the Watermark are generally suitable for proving ownership or authentication. Therefore, we established a framework based on asymmetric public-key cryptography which is used for exhaustive authentication with the help of *Blind Watermarking* techniques.
In addition to the traditional analogue, and newer digital, radio and television transmission means, programme broadcasts may also be received over the Internet. The *broadcaster* (BC) is an entity which broadcasts content for general consumption over their assigned channels, for instance an IP network (IPTV). In many jurisdictions broadcasters have regulatory obligations which attempt to protect the intellectual property [5] and copyrights of authors, songwriters, performers, actors, publishers, etc. Furthermore, in some jurisdictions there exists bodies charged with the defense of the rights of intellectual property and copyright holders and the calculation, charging and collection of performance royalties on the use of these protected works. Currently, there are several cases in which broadcasters cannot confidentially confirm that their royalties liabilities are correctly calculated. This is because they currently do not employ a viable automated system to measure what protected works are broadcasted, how often and when. Therefore a gap has opened up in the actual amount charged by the rights bodies and the correct payable royalties liability of the broadcaster.
This paper focuses on methods and procedures to close this gap and to support means for obtaining more detailed information about streamed content. A framework for authentication based on PKI for the parties involved is introduced, as is a framework for Watermarking rich media content. The main objective of these frameworks is to provide the ability to the broadcaster to be able to prove the amount of streamed media actually used to the agency responsible for monitoring broadcasters. This agency is called the *monitoring agency* (MA) and is

*VICOMTech Research Center, Paseo Mikeletegi 57, E-20006 San Sebastian, Spain
†Horst Görtz Institute for IT Security, Ruhr-University Bochum, Gebäude IC 4, D-44780 Bochum, Germany

charged with supervising broadcasting stations for purposes of calculating the royalties to be paid by broadcasters to the rights body. Monitoring agencies are often instructed by a *rights entity* (EX), which is a competent organisation charged with protecting the rights of intellectual property and copyright holders, and negotiating recompensation for the exploitation of these works.
However, this framework is in the interest of each involved party and attempts to address the needs of all parties. The broadcaster can be sure that it is being charged fairly. The monitoring agency can offer services and get paid for doing so. The rights entity can perhaps calculate that a broadcaster should pay more, or less, royalties than they are actually paying. For each party a substantial advantage can be seen by implementing such a framework.

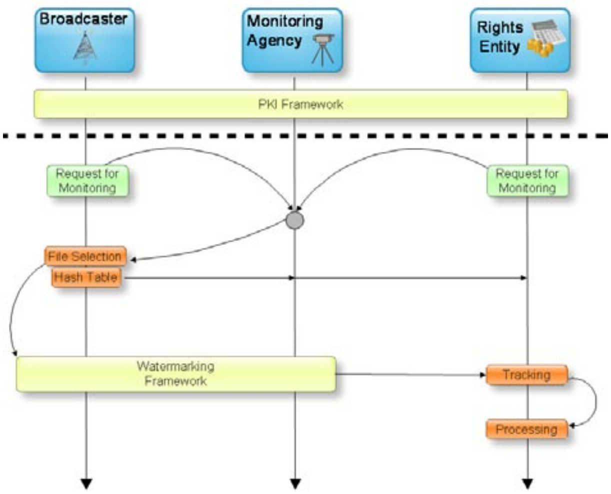


Figure 1: General Framework Overview

2 Framework Overview

A general overview of the framework with its three parties can be seen in Figure 1. It makes use of two additional frameworks, the PKI and the Watermarking Framework. The PKI Framework (section 4), is used for establishing a trust network between all of the involved entities. It makes use of a root Certificate Authority (CA) in which each participating entity must trust.

In order to establish a working PKI, each of the parties has to process one or more protocol steps. The PKI establishment has the purpose of distributing authenticated private and public keys utilising Digital Certificates. To start the process of monitoring, a "Request for Monitoring" is sent to the monitoring agency. Afterwards, broadcaster selects a piece of content which he wants to stream and computes the corresponding hash table. This hash table is carried over a secure and authenticated channel to the MA as well as to the EX. Subsequently, the broadcaster initiates the process defined by the Watermarking Framework.
The Watermarking Framework (see section 5) specifies procedures for Watermark embedding, retrieval and verification of Watermarks in media streams. The broadcaster will sign the stream which is about to be broadcasted with his private key. Then the corresponding signature is embedded into the media stream with a known Watermarking technique. Further on in the process, the monitoring agency will extract the Watermark and verify the signature. Therefore, the agency may be sure that only the original broadcaster broadcasted the media content, due to the fact that additional security metadata, such as timestamps and identifiers, are used. Additionally, EX can also verify the signature in order to prevent abuse by the MA.
The objective of the whole framework is to let the broadcaster mark the file stream uniquely but also provides the monitoring agency with the possibility to identify the broadcast stream and therefore the corresponding broadcaster. In this paper we focus on the novel IPTV broadcasting infrastructure [2]. However, our framework should be applicable to any broadcasting infrastructure irrespective of the underlying distribution network.

3 IPTV Overview

IPTV is a new method of delivering digital video and audio content across IP networks [10]. Today, the industry is going through a profound transition, migrating from conventional TV to the era of digital technology. IPTV technology provides several advantages but service providers face a set of specific barriers, namely the availability of sufficient network capacity, especially in the last-mile or local loop section of the broadband network that lies between core telecommunications network, or back-haul trunks, and the end-user. There are currently six principal types of broadband access networks which have sufficient capacity and scalably to meet the bandwidth requirements of IPTV:

- Optical Fibre
- Digital Subscriber Line (DSL)
- Cable TV
- Satellite Broadband
- Fixed Wireless Broadband
- 3G Mobile Data Networks

Several different applications can be provided by IPTV service providers over a broadband connection, however the two key IPTV applications typically deployed by service providers are broadcast digital TV and content on demand (CoD). In this paper we only focus on digital TV broadcast.

Security is the number one priority for IPTV service providers because video content producers are reluctant to grant license rights to distribute premium content over digital networks unless there are effective mechanisms in place, which will secure that content. There are a number of IPTV content protection schemes available on the market. These protection schemes fall into two broad categories: Conditional Access (CA) and Digital Rights Management (DRM) environments. The CA system is primarily responsible for ensuring unauthorised access of the IPTV service, while the DRM system enforces content owners business models and granted usage rights.

4 PKI Framework

Trust forms the basis of all communication, be it physical or electronic. In the case of electronic communication, building trust is quite difficult as the identity of the other entity remains concealed. In our specific case, trust is also the foundation for the calculation and collection of royalties. The three entities, the rights entity, the broadcaster and the monitoring agency, need to trust the CA. Therefore, a PKI needs to be established which provides procedures to generate, distribute, and utilise keys and Certificates and so helps to build up a trust relationship.
We propose a single-CA architecture which makes use of a superior independent Certificate Authority. The CA must not be involved or integrated in metering and should be independent and impartial. Within the single-CA architecture, all entities trust the CA and therefore can validate and verify each others Certificates and then communicate. In later communications, the CA need not be involved.
After a successful PKI-establishment, the broadcasting entity could sign a message and send it to the monitoring agency or indeed to the rights entity and both entities could be assured, that the message was sent by the broadcaster.

5 Watermarking Framework

The Watermarking Framework specifies the communication protocol between the broadcaster and the monitoring agency in which the rights entity is not involved. Furthermore, the Watermarking Framework provides a detailed insight into procedures for creating, detecting, extracting and verifying the Watermark.

5.1 Overview

The chief characteristic of a traditional Watermarking scheme for copyright protection, or DRM, is that the Watermark cannot be separated from the medium without knowledge of a secret value. We, in our specific case, target on another characteristic: sender authentication. It should be possible to identify the broadcasting station unambiguously and show exactly who broadcast what stream and when.
Therefore, our Watermark information contains a digital signature issued by the broadcaster that definitively identifies the broadcaster. Each entity that receives the broadcast stream and owns the corresponding broadcaster Certificate, can clearly verify the distributed stream with the help of the corresponding PK.

5.2 Signature Schemes

A principal requirement to all Watermarking systems is the need for a small Watermark. The larger the Watermark, the larger are the chances for adversely affecting the quality of the streamed media. Therefore, the signature scheme output has to be as small [11] as is possible to be able to embed the Watermark as often as possible and to be repeated multiple times throughout the stream. While the typical RSA 1024-bit signature output is large, several alternative schemes were researched.
The Nyberg-Rueppel ([12], hereafter NR) signature scheme focuses on the size of the input and output and is a DSA-like signature with message recovery. The drawback of this signature type, the fixed length of input, is not given in our case because the message m (see (1)), which is used for exact identification of the stream, is always brought to a fixed length through a given hash function. NR is perfectly suited to messages shorter than ten bytes but leaves the question of dealing with short messages, of say fifteen bytes, unanswered. In our specific case, the hash to be signed is exactly 10 bytes long and brings only a marginal risk of collision. Message recovery [13], another characteristic of NR signatures, provides means so that the original message can be extracted out of the signature.

5.2.1 Short Hash Methods

Hash functions are often used in digital signature algorithms. The message m that is about to be hashed, in our case, consists of an identifier string *ID-str* concatenated with an ID number *ID-num* and an unique times-tamp *ID-time*:

$$m = \textit{ID-str} \mid\mid \textit{ID-num} \mid\mid \textit{ID-time} \tag{1}$$

The ID-str could be represented through the name of the media content, for instance. The ID-num could be an identification number. The ID-time is a unique timestamp which prevents replay-attacks. This means, that an adversary may not record the stream and broadcast it later again on an authorised channel which is also monitored.

5.2.2 Hash Table for Verification Purposes

A hash table *ht* in our specific case is a data structure that associates the hash value with ID-str, ID-num and ID-time. The hash table contains several important attributes and is essential for the verification process by the MA.
Transferring the hash table *ht* to the MA, can be compared to the cryptographical *commitment scheme*, visualized in Algorithm 1.

Algorithm 1 Secure and Authentic Hash Table Distribution

Summary: during the *commitment phase*, the hash table is transferred to MA and EX. During the *opening phase*, BC proves to MA and EX that he is broadcasting one of the items in the hash table.

1. ***commitment phase***:

   1. BC $\longrightarrow$ MA: $\text{enc}_{PK_{MA}}(\text{sign}_{SK_{BC}}(ht))$
   2. MA $\longrightarrow$ EX: $\text{enc}_{PK_{EX}}(\text{sign}_{SK_{BC}}(ht))$

2. ***opening phase***:

   1. BC $\longrightarrow$ MA: watermark($\text{sign}_{SK_{BC}}(hs)$)
   2. MA: extract *signature* from stream with the help of beacon
   3. MA $\longrightarrow$ EX: $\text{enc}_{PK_{EX}}(\text{sign}_{SK_{BC}}(hs))$

The prover, respectively the BC, sends the "commitment" in form of the hash table *ht* to the verifier (the MA). MA will forward the signed hash table to the rights entity but encrypts it with the corresponding PK_{EX} in order to guarantee secrecy which is needed to prevent other parties from viewing the hash table. This can be seen as the *commitment phase* and takes place directly after having chosen the file to be streamed. The encryption is necessary due to the possibility that the hash table of a potential business rival might be seen by another party. Later, after broadcasting the media content, the verifier can scrutinise, with the help of the message recovery characteristic of the signature, whether the BC broadcast the content correctly or not (*opening phase*).

5.2.3 Case Study: Video Broadcaster

The *Internet Movie Database (IMDB)* [1] published interfaces for several systems to access the IMDB locally. For our case study, we downloaded the complete IMDB title textfile which contains currently 1.206.730 different movie titles. We used the movie title as a ID-str and created a unique number used as the ID-num. The time-stamp ID-time was the current date parceled as a unixtimestamp. An example assignment between unixtimestamp and normal time can be seen in (2).

$$05/07/1982@00:00 \Longrightarrow 389592000 \tag{2}$$

For instance, in our simulation, m looked like this:

$$\text{m} = \underbrace{\text{Title A}}_{\textbf{ID-str}} \quad || \quad \underbrace{23754}_{\textbf{ID-num}} \quad || \quad \underbrace{534056}_{\textbf{ID-time}} \tag{3}$$

We created 1,206,730 different messages m and subsequently hashed them with MD5 and SHA-1. Afterwards, we extracted the first 10 bytes which satisfy the first 20 characters of the output HEX value. No collisions were detected for both hash functions, MD5 and SHA-1, even with only using the first 10 bytes of the hash-sum.

$$hs = [0...9]hash(m) \tag{4}$$

Finally, a theoretical possibility to create a 2nd-preimage attack on our used short hash methods remains. Because of the reduced length of the hash value, our methods don't have the complete potential strength of 2^{160} (SHA-1) respectively 2^{128} (MD5) 2nd-preimage resistance. The search space would be reduced to 2^{80} respectively according to the birthday paradox on 2^{40}.

5.3 Suitable Watermarking Algorithms

In our specific case, the Watermark should have special control characteristics which are required to guarantee the ability to verify the embedded signature by the monitoring agency.
Spread-spectrum [14] technologies establish secrecy of communication by performing modulation according to a secret key in the channel encoder and decoder. Our specific scenario does not focus on secrecy but on authentication. Therefore, the beacon used for encoding and decoding, only contains the information how to process these steps. The beacon is not a secret value.

5.3.1 Proposed Watermarking Algorithm

Basically, a Watermarking system for our purposes can be described by a tuple $\langle \mathcal{O}, \mathcal{S}, \mathcal{W}, \mathcal{H}, \mathcal{P}, \mathcal{G}, C_S, E_H, D_H, V_P \rangle$ where $\mathcal{O}$ is the set of all original data, a video stream for instance. The set $\mathcal{S}$ contains all secret keys needed for creating an unforgeable signature. $\mathcal{W}$ represents the set of all Watermarks (signatures, in our case) and $\mathcal{H}$ the

[1] http://www.imdb.com

set of all beacons. Beacons in our scenario are markers that signify the presence and start of a Watermark bit sequence in the signal. The beacon substitutes the key in normal Watermarking systems. $\mathcal{P}$ describes the set of public keys which are needed to verify the signature and $\mathcal{G}$ represents the set of Certificates issued by the CA.
Four functions are described as followed:

$$C_S : \mathcal{O} \times \mathcal{S} \longrightarrow \mathcal{O} \tag{5}$$
$$E_H : \mathcal{O} \times \mathcal{S} \times \mathcal{W} \times \mathcal{H} \longrightarrow \mathcal{O} \tag{6}$$
$$D_H : \mathcal{O} \times \mathcal{H} \longrightarrow \mathcal{W} \tag{7}$$
$$V_P : \mathcal{W} \times \mathcal{P} \times \mathcal{G} \longrightarrow \{1,0\} \tag{8}$$

C_S focuses on creating the corresponding Watermark through a signature. E_H describes the function for embedding the Watermark and D_H respectively the function for extracting it. Furthermore, V_P stands for the verification function needed to check if the Watermark is valid. The Watermark w is created with

$$w = sign_{SK_{BC}}(hs) \tag{9}$$

and outputs a short bit-string which contains the signature of the reduced hash-sum. See (4) for further details about the reduced hash-sum hs.

5.4 Embedding the Watermark

In this subsection we focus on the embedding process of the signature/ Watermark. Hartung and Girod proposed in 1998 [15] a method which focuses on Watermarking MPEG-2 video data. We adopt the proposed methods for our purposes of embedding the signature into a given video broadcast stream. For further information, the interested reader is referred to [15].

5.5 Retrieval of the Watermark

The proposed methods rely on *Blind Watermarking* techniques and therefore do not need the original video stream in the retrieval process. For further information, the interested reader is referred to [15].

5.6 Verifying the Signature

It is possible for the monitoring agency to verify the signature which is represented by the extracted bit sequence. The method V_P verifies the signature with the help of the corresponding public key and Certificate. The used public key for verifying is taken from the Certificate in order to be sure, that only the public key belonging to the correct broadcaster is used.

6 Conclusions and Future Work

The schemes proposed in this paper may be viewed as attractive to both broadcasters and rights agencies. This model provides the broadcaster and the rights entity with an automated and trust worthy method for measuring the exploitation of protected works. The paper introduces the concept of an independent third party that monitors and balances the interests of the broadcaster and rights entity. We discuss the rapidly evolving technologies

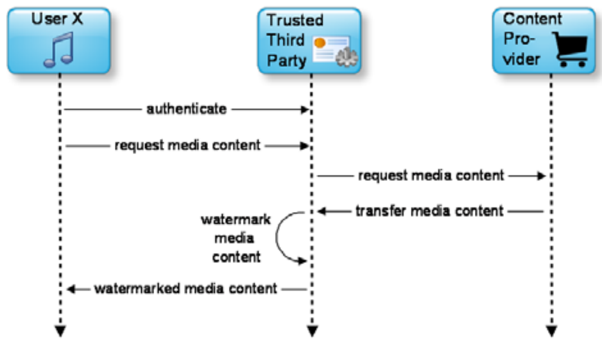


Figure 2: Abstract Proposal for DRM-substituting Business Model

and distribution models faced by the entertainment and broadcasting sectors. Then we discuss next generation media distribution using IPTV as an example. We evaluate established short signature schemes, such as Nyberg-Rueppel, that could be integrated into a final system.
Our model could function as a compliment, or an alternative, to established DRM models.
Therefore, in Figure 2 we propose an exemplified scheme which could substitute current DRM models. If a user wants to buy a media content (audio, video or image content) from a content distributor, the TTP handles the whole process. The request for the specific media content gets proxied by the TTP for providing anonymity. Afterwards, the content is Watermarked by the TTP with a user-specific signature and sent back to the user. This means, that the content provider will never get the knowledge of the user's secret key.
Clearly, Watermarking has a number of characteristics that make it an ideal technology for enabling a variety of media distribution business models. However, historically robustness has been a chief weakness of Digital Watermarking techniques.
A variant of our model could be used by existing online music services to modify their current DRM schemes toward an intellectual property preserving framework based on personalised Watermarks. Digital Watermarking schemes present an alternative to regulatory measures. Although not covered in this paper, the current body of national and European law provides legal protection for Watermarking and Digital Certificate technologies. A robust Digital Watermark can jump the analogue hole. This might mitigate the need for Broadcast Flags, TPM like chipsets or signal degradation on playback device to be mandated by law.
State of the art Watermarking techniques have taken substantial steps forward in addressing the issue of robust-

ness. Currently, the big four record labels rely on modern robust Watermarking algorithms to sell DRM-Free MP3 files through the Amazon MP3 store. There is a direct linear relationship between the robustness of a Watermark and the size of its payload. High definition content presents the ideal conditions to improve Watermark robustness. It has a greater than linear increase in size over standard definition. Therefore there is a greater quantity of available data in the signal to embed a complex and robust Watermark.
A Digital Certificate can be used to enter into a contract. A media file Digitally Watermarked with a value derived form a Digital Certificate may be viewed as a type of Smart Contract. This provides the distributor with a means to trace the file to the purchaser, should it appear on P2P networks. More importantly, the act of signing the media file motivates the consumer not to make an unauthorised copy of the file. Ideally the incentive to the consumer would be lower prices. The benefit to the distributor would be increased sales due to reduced piracy.

References

[1] Gerard Driscoll. *Next Generation IPTV Services and Technologies.* Wiley-Interscience, 2008.

[2] Neill Wilkinson. *Next Generation Network Services*, volume 1. Wiley & Sons, 2002.

[3] M Peitz and P Waelbroeck. Piracy of digital products: A critical review of the economics literature. cesifo working paper series no. *Information Economics and Policy*, (1071):2003, 2003.

[4] Ingemar J. Coxy, Joe Kiliany, Tom Leightonz, and Talal Shamoony. A secure, robust watermark for multimedia, 1996.

[5] Chun-Shien Lu. *Multimedia Security: Steganography and Digital Watermarking Techniques for Protection of Intellectual Property.* Idea Group Publishing, 2005.

[6] W J Dowling and F M Boland. Watermarking digital images for copyright protection. pages 250–256, 1996.

[7] D Kirovski, H Malvar, and Y Yacobi. A dual watermarking and fingerprinting system. Technical report, ACM Multimedia, 2001.

[8] Jian Zhao. Applying digital watermarking techniques to online multimedia commerce. In *In Proc. of the International Conference on Imaging Science, Systems, and Applications, Las Vegas*, 1997.

[9] Zbigniew Kotulski Wojciech Mazurczyk. Covert channel for improving voip security. In *Advances in Information Processing and Protection*, 2007.

[10] Nigel Seel. *Business Strategies for Next-Generation Networks.* Auerbach Publications, 2007.

[11] David Naccache and Jacques Stern. Signing on a postcard. *Lecture Notes in Computer Science*, 1962:121–??, 2001.

[12] Kaisa Nyberg and Rainer A. Rueppel. A new signature scheme based on the dsa giving message recovery. In *Proceedings of the 1st ACM CCCS*, Fairfax, 1993. ACM.

[13] G. Ateniese and B. de Medeiros. A signature scheme with message recovery as secure as discrete logarithm. 1999.

[14] Ingemar Cox, Joe Kilian, Tom Leighton, and Talal Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing*, 6(12):1673–1687, 1997.

[15] Frank Hartung and Bernd Girod. Watermarking of uncompressed and compressed video. *Signal Processing*, 66(3):283–301, 1998.

Workshop on Digital Preservation Weaving Factory for Analogue Audio Collections

Organized by

Fondazione Rinascimento Digitale (Italy), University of Florence (Italy)

Coordinated by

Prof. Francesco Carreras (ISTI-CNR, Italy)

Emanuele Bellini (Fondazione Rinascimento Digitale, Italy)

AXMEDIS 2008: proceedings of the 4th International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution. Workshops, Panels and Industrial Applications. Florence, Italy 17-19 November 2008
edited by Paolo Nesi, Kia Ng, Jaime Delgado
ISBN: 978-88-8453-811-6 (online). ISBN 978-88-8453-810-9 (print).

From Analogue to Digital: the cycle of digital content production from sound documents of historical archives

Alberto Gaetti - MART^{Lab}

Abstract—**The following document focuses on the cycle of digital content production, showing basic aspects of the workflow developed at MART^{Lab}: inspection and cataloguing of sound documents, carrier restoration, digitization, storage and preservation of original carriers and their digital surrogates, access, audio restoration and production of sound documents are the main tasks described in this article.**

Index Terms—**Audio Heritage, Audio Restoration, Digitization, Preservation, Storage.**

I. Introduction

DIGITIZATION of sound documents and good conservation practices of original carriers have become the preferred way to ensure preservation of sound documents and access to their contents, reducing at the same time the wearing of original carriers due to repeated playback.

The process that brings from the original sound document, generally an analog recording (although today incoming problems of conservation of the so called *born-digital* must be also considered) to the digital surrogate must be carefully planned, in order to guarantee the best possible:

- integrity of the original carrier;
- integral and most accurate transfer of audio content;
- retrieval of non audio, ancillary and secondary information contents.

At every step of the process, the production of metadata about the work (i.e. description of the carriers, equipment parameters, operators choices, etc.) is mandatory. Care must be taken in thoroughness and coherence of the attribution of such metadata to the digital surrogate: there is a subsequent need for the use of persistent Unique Source Identifier applied to digitized items and their digital surrogates.

The importance of the management of the entire process, that means definition of the policies that guide the strategy of every process, administration and control of the tasks that involve audio carriers, data and metadata, should not be underestimated.

The Author is with MART^{Lab} Laboratorio di Musica e Audio, Ricerca, Recupero, Restauro e Tecnologie – c/o Conservatorio di Musica "L. Cherubini", Piazza delle Belle Arti, 2 – 50122 Firenze, Italy (Alberto Gaetti, e-mail: alberto.gaetti@martlab.it).

The process of treatment of sound documents starts when a collection or a historical archive decide for digitization. The principles that inspire this decision draw the path of the entire process of production of digital surrogates.

II. The MART^{Lab} Workflow

There are many ways to plan and manage a process of treatment and digitization of sound documents [1-5]; at MART^{Lab} we developed a protocol of treatment, digitization and production of sound documents based on IASA Guidelines [6] [7] that took advantage of the experience of PrestoSpace [8] and CASPAR [9] projects. The MART^{Lab} workflow can be synthesized in five Functional Areas:

- administration&control
- inspection/cataloguing
- carrier restoration/digitization
- storage/preservation
- access/audio restoration&production

Every Functional Area is composed of one or two Sections. Each section gathers many elementary stages or tasks that operate on sound documents, data and metadata.

Basically each Section presents inputs and outputs that interact with other sections; inputs and outputs may be represented by analog and digital carriers, analog and digital data, ancillary information, metadata, administrative instructions, etc.

The following document focuses on the cycle of digital content production, showing basic aspects of the workflow developed at MART^{Lab}.

A. Administration & Control Unit

Every Section reports its results to the Administration & Control Unit (ACU) and waits for instructions to proceed with its work.

Management is demanded to the ACU which performs the following tasks:

- determine the principles and the policies that guide each step of the process of digital surrogates production;
- collect and analyze reports and results (metadata) of the different Sections, in order to define priorities for the treatment of sound documents and to manage resources.

B. Inspection/Cataloguing Functional Area

Inspection Section

Inspection of the collection should be made both in the original storage location and in the laboratory of digitization, after the delivery of the items.

In the original storage archive the inspector should investigate:

- provenance of the collection;
- permanence of the collection in the location;
- number of items of the collection;
- historical information about the collection;
- general conservation conditions of the collection;
- general information about the collection.

After the delivery of the items to the temporary storage location of the laboratory, the inspector can focus his attention on every item and investigate:

- conservation condition of the item;
- materials and manufacturing of the item;
- authorship of the item (who recorded it, how and when; the real content recorded on the item);
- historical and secondary information about the item;
- general information about the item.

It should be noted that some kind of information can be retrieved only during the playback of the carrier (i.e. the actual content recorded on the item, playback parameters, etc.), so the task of inspection of one item starts at its delivery and effectively ends after its playback is finished.

Cataloguing Section

Although different archives have their own system of cataloguing, the laboratory should label each item with a Unique Source Identifier – USID – that will link the original sound document to its digital surrogate and metadata.

Every task of the Inspection/Cataloguing Functional Area produces metadata that should be collected and reported to the ACU. ACU main tasks for the Inspection and Cataloguing Sections are:

- transfer of the collection to the temporary archive of the laboratory.
- definition of temporary conservation conditions.
- definition of USID syntax.
- definition of item digitization priority of the items.
- planning time and resources for carrier restoration, digitization and production.

C. Carrier Restoration/Digitization Functional Area

Carrier Restoration Section

Each item should be treated in order to obtain the best transfer of its content; main tasks for this Section are:

- deep analysis of defects, materials, manufacturing and conservation conditions;
- remove contamination, dust, and traces of glue and stickers (polishing/washing);
- remove mechanical deformations in discs and tapes (heat treatment);
- remove spooling stress and print-through in tapes (re-spooling);
- reduce/remove hydrolysis artifacts (heat treatment);
- specific audio carriers treatments and restoration.

Digitization Section

For playback and A/D conversion the following tasks have to be performed:

- playback equipment must be calibrated to standard parameters;
- each item must be analyzed for the definition of the recording format and the setup of the playback parameters (speed, equalization, mechanics, etc.) in order to obtain the best retrieval of the audio content;
- digitization equipment must be calibrated and set to standard parameters;
- temporary storage of digital data;
- carrier preparation for long term conservation.

Every task of the Carrier Restoration/Digitization Functional Area produces metadata that should be collected and reported to the ACU. ACU main tasks for the Carrier Restoration and Digitization Sections are:

- extraction of a specific item from the temporary archive in the laboratory on the basis of the planned digitization priority;
- analog playback equipment calibration (based on items formats, number of working hours of the equipment, wearing, etc.);
- digitization equipment calibration;
- definition of digitization parameters (bit depth, sampling frequency, etc.);
- definition of digital file format;
- data and metadata integration;
- data and metadata migration to the storage area;
- transfer of the digitized item to the temporary archive of the laboratory.

D. Storage/Preservation Functional Area

Storage Section

In this Section, operators manage the storage of digital data and metadata, and take care of temporary and long term conservation of original carriers. Main tasks are:

- temporary conservation for digitized items in the laboratory archive;
- transfer of data and metadata from the digitization area to the storage area;
- cloning of data and metadata (redundancy);
- migration and refreshing of data and metadata;
- system performance check & fault detection;
- access management.

As a matter of fact the Storage Section should be planned and activated once items have been delivered to the laboratory, just after the first inspection of the collection. On the basis of the strategy recommended by the Preservation Section, Storage Section's operators should provide support and solutions for the sound documents lifespan storage.

Preservation Section

Preservation is not an operational stage effectively, rather a set of strategies and practices related both to conservation and access to items and their contents. From this point of view Preservation manages legacy sound documents, their digital surrogates and metadata in a complex of relationship facing technological developments, formats obsolescence, conservation issues, access systems and solutions, etc.

In this workflow, Preservation is located between the Storage Section and the Access Section in order to emphasize that the approach to digital content production from historical sound documents have to grant enduring access to contents and their related metadata. The Preservation Section prove:

- strategies for digital data and metadata integration and relationship;
- strategies and schedule for migration and refreshing of digital data and metadata;
- software and hardware storage update and upgrade;
- strategy and good practices of conservation, test and control of original sound documents.

Every task of the Storage/Preservation Functional Area produces metadata that should be collected and reported to the ACU. ACU main tasks for the Storage and Preservation Sections are:

- definition of storage conditions for temporary and long term conservation of digitized items;
- definition of long term preservation policy for digitized items;
- definition of data and metadata formats;
- definition of the structure of the digital mass storage system;
- definition of the policy of software and hardware storage solutions update and upgrade;
- definition of the policy of data and metadata migration and refreshing.

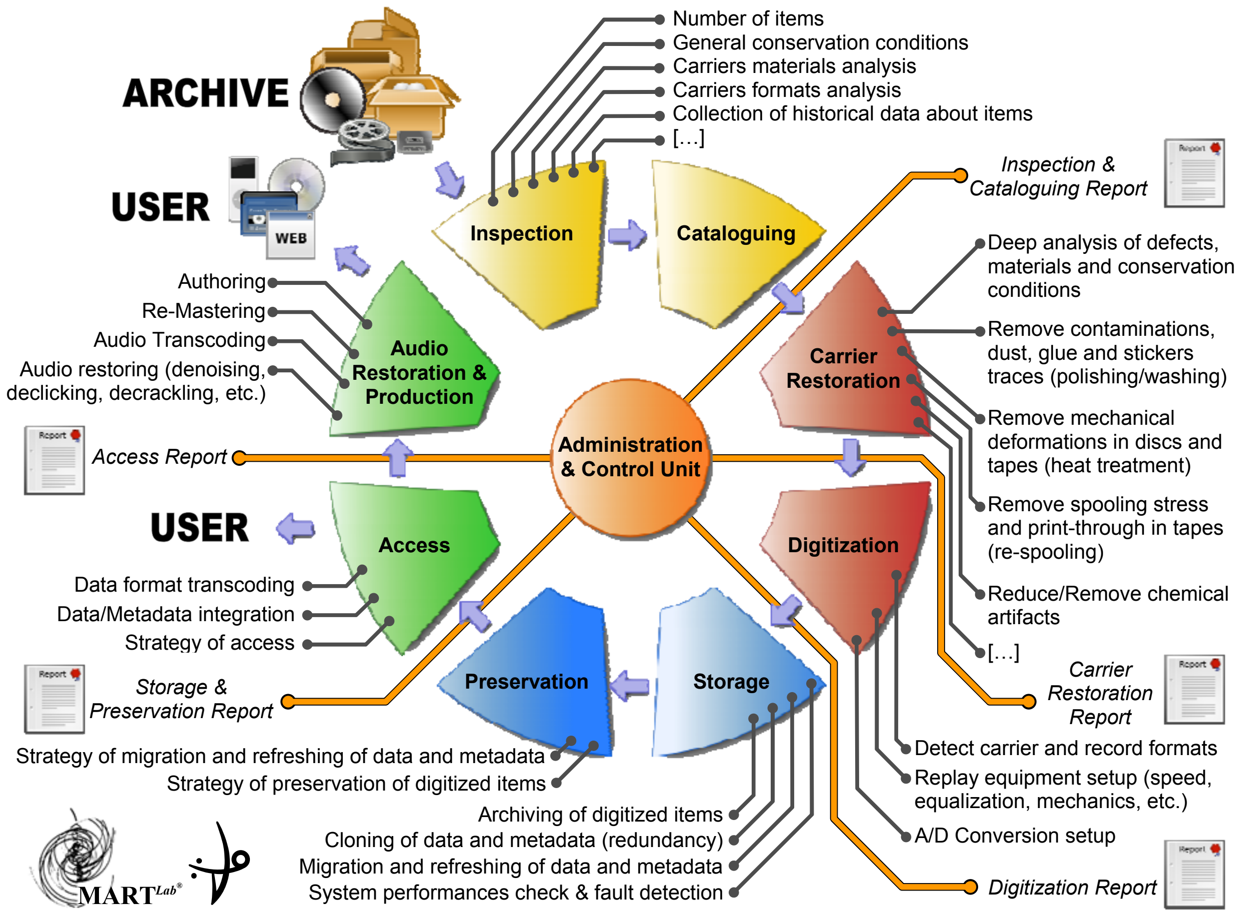


Fig. 1 – Diagram of the cycle of digital content production from historical sound documents.

E. Access/Audio Restoration & Production Functional Area

Access Section

Access to digital surrogates must be restricted to authorized users. Open access should be provided to copies/clones of digital data and to selected set of metadata. The choice of data and metadata formats must be aimed at the access media/system. The main tasks of this section are:

- development of a strategy of access;
- data format transcoding;
- metadata format transcoding;
- data and metadata integration
- provide the access system to contents.

Access to original items must be strictly restricted to operators and submitted to preservation plan.

Audio Restoring & Production Section

In order to produce digital content, audio restoring and re-authoring of digital surrogates have to be achieved. The main tasks of this section are:

- audio restoring (denoising, declicking, decrackling, etc.);
- data and metadata format transcoding;
- re-mastering;
- authoring/production;
- distribution.

Every task of the Access/Audio Restoration & Production Functional Area produces metadata that should be collected and reported to the ACU. ACU main tasks for the Access and Audio Restoration & Production Sections are:

- management of copyright and ownership;
- definition of the policy of access;
- definition of the final carriers for the production;
- definition of the level of data and metadata integration;
- definition of the policy of data quality format transcoding;
- definition of the policy of audio restoring and re-mastering.

III. Massive digitization

The development of this protocol and workflow was possible thanks to the experience of MART^{Lab} in digitizing items of small collections of non-homogenous sound documents (carriers of different materials, formats and recording parameters, different conservation conditions, etc.).

Under these conditions it is very difficult to undertake a massive digitization (i.e. sound documents parallel digitization): each item shows its own peculiar characteristics and needs the complete attention of the operator during the entire process.

In order to plan a massive digitization for a large-scale collection some conditions have to be respected; the collection must be composed of:

- items that present almost the same conservation conditions (same rate of physical or chemical degradation);
- items of few physical formats;
- items recorded with almost the same formats and parameters;
- items recorded with well-known equipment.

Moreover the playback equipment for the massive digitization must be homogenous, well calibrated and provided with some kind of automation.

Digital equipment should be set for massive digitization, providing multiple channel A/D conversion and asynchronous playback and recording facilities. Furthermore proper signal routing should allows accurate audio monitoring.

In massive digitization care must be taken in the thoroughness of the attribution of metadata to the proper digital surrogate at each stage of digitization.

IV. Conclusion

The workflow for digital content production from historical sound documents showed in this article is a synthesis of the MART^{Lab} digitization protocol.

The MART^{Lab} protocol for the production of digital surrogates and metadata from original sound documents has been developed thanks to the experience in digitization of sound documents of small collections of non-homogenous items and is based on main standards and guidelines requirements. Inspection and cataloguing of items, carrier restoration, digitization, storage and preservation of original carriers and their digital surrogates, access, audio restoration and production of sound documents was the main tasks described in this article.

References

[1] Colorado Digitization Program - Digital Audio Working Group, “Digital Audio Best Practices”, Version 2.1, October 2006. [Online] http://www.bcr.org/cdp/best/digital-audio-bp.pdf

[2] J. Williams, S. Paske, S. Dast, “Audio Procedures and Workflow for the University of Wisconsin Digital Collections Center (UWDCC)”, University of Wisconsin, 2004. [Online] http://uwdcc.library.wisc.edu/documents/AudioWorkflow.pdf

[3] Indiana University Digital Library Program, “Sound Directions: Best Practices for Audio Preservation”, April 2008, [Online] http://www.dlib.indiana.edu/projects/sounddirections/papersPresent/index.shtml

[4] J. Henriksson , N. Wallaszkovits, “Digitisation workflow for analogue open reel tapes”, TAPE - Training for Audiovisual Preservation in Europe. [Online] http://www.jazzpoparkisto.net/audio/

[5] Cornell University Library, Digital Preservation Management, 2007. [Online] http://www.icpsr.umich.edu/dpm/

[6] IASA Technical Committee - Standards, Recommended Practices and Strategies, “The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy”, IASA-TC 03 Version 3, December 2005.

[7] IASA Technical Committee - Standards, Recommended Practices and Strategies “Guidelines on the Production and Preservation of Digital Objects”, IASA-TC 04, August 2004. [Also available in Italian as “Linee guida per la produzione e la preservazione di oggetti audio digitali”, AIB, 2007]

[8] Presto Space - Preservation towards storage and access. Standardised Practices for Audiovisual Contents in Europe. [Online] http://www.prestospace.org

[9] CASPAR - Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval. [Online] http://www.casparpreserves.eu

IASA TC 03 and TC 04: Standards Related to the Long Term Preservation and Digitisation of Sound Recordings

Nadja Wallaszkovits
Phonogrammarchiv, Austrian Academy of Sciences
Nadja.Wallaszkovits@oeaw.ac.at

Abstract

The paper introduces the aim and the principles of the current standards in audio archiving "IASA-TC 03: The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy" *and* "IASA-TC 04: Guidelines on the Production and Preservation of Digital Audio Objects". *The paper provides an overview over the actual solutions adopted for repositories concerning the main topics that audiovisual archives are confronted with, such as basic ethical decisions, obsolescence of formats, principles of safeguarding the information, selection of best copy and carrier restoration, optimal signal retrieval from original carriers, unmodified transfer to a new target format, digital target formats and resolution, digital mass storage systems and small scale approaches. Additionally the paper will outline the main basic strategic considerations by means of practical application in small scale digitization projects.*

1. Introduction

Focusing the current situation, by far the majority of audiovisual documents worldwide still exists in analogue representation. Therefore, in an increasingly digital environment, the safeguarding of audiovisual cultural heritage has become a more and more challenging task.

Digital audio has, over the past few years, reached a level of development that makes it both effective and affordable for use in the preservation of audio collections of every magnitude. The integration of audio into data systems, the development of appropriate standards, and the wide acceptance of digital audio delivery mechanisms have replaced all other media to such an extent that there is little choice for sound preservation except digital storage approaches. Digital technology offers the potential to provide an approach that addresses many of the concerns of the archiving community through lossless cloning of audio data through time.

The transfer of archive holdings into the digital domain is a major topic that needs an amount of strategic considerations. The processes of converting analogue audio to digital, transferring to storage systems, of managing and maintaining the audio data, providing access and ensuring the integrity of the stored information, present a new range of risks that must be managed to ensure that the benefits of digital preservation and archiving are realised. Failure to manage these risks appropriately may result in significant loss of data, value and even audio content.

2. IASA TC03: The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy

2.1 Background and Intention

IASA-TC 03 *"The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy" http://www.iasa-web.org/downloads/publications/TC03_English.pdf* has been edited by Dietrich Schüller. Contributors are members of the IASA Technical Committee, in detail George Boston, George Brock-Nannestad, Lars Gaustad, Albrecht Häfner, Dietrich Schüller and Tommy Sjöberg, and the publication is reviewed by the IASA Technical Committee. The first version has been published in February 1997. At that time digital archiving - mainly on tape based media - was already widespread, but analogue archival masters were still recommended as adequate medium. In version 1 the core principles of digital archiving were already defined.

In September 2001 the second edition was presented, wherein the viability of digital archiving was unanimously accepted. Besides a major rearrangement of contents the guidelines became more practically oriented.

The second revision version 3, released in December 2005, shows a closer connection to IASA-TC 04 (the more practically oriented part of these guidelines, published in 2004). While TC 03 outlines the principles, TC 04 has become a practical reader to support daily digitisation work. Consequently, practical matters have been eliminated from TC 03, and the sequence of issues has been aligned to match with TC 04.

IASA-TC 03 *"The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy"* aims to identify problem areas and to propose recommended practices for use by sound and AV archives in today's technical environment. These recommendations are intended to help the reader to focus on the various issues relating to responsible audio archiving practice a balance between the ideal situation and the real world. The major aim of this publication is to reach consensus amongst preservation specialists and to spread consensus amongst audiovisual archivists.

At the same time it uses a consistent terminology and may be read by people with financial responsibility for a collection as well as by technically trained staff.

2.2 Basic Essentials of IASA TC03

The publication starts with a brief discussion of ethical considerations, outlining that *"TC 03 is not a Code of Ethics for all aspects of sound archiving. It covers, however, the ethical consequences resulting from the technical aspects of recording, preserving and accessing sound documents within the framework of the technical development offered by today's market situation" [1].*

Chapter 1 gives a definition of the four basic tasks of sound archives, which are outlined as acquisition, documentation, access and preservation. To fulfil this list of requests, sound archives have to ensure that the information placed in the care of the collection is preserved and that the integrity of the information is guaranteed. This is from today's point of view only possible in the digital domain, provided that the information to be preserved is digitised accurately and without any loss of quality.

This claim leads directly to chapter 2. Herein it is outlined that audiovisual carriers contain primary and secondary information: primary information consists of the sonic content, the essence that was intended to be recorded. Secondary information includes all metadata and associated materials and information, as well as the technical representation. All kinds of information are part of the document and have to be preserved. This rises the problem that in conventional transfer processes some technical information is lost, e.g. the high frequency bias frequency of a magnetic tape, which could serve as reference for irregular speed deviation of recordings on analogue magnetic tape [2].

The instability and vulnerability of audio carriers is covered by the following chapter 3. Herein it is explained that all audio-visual carriers are instable and vulnerable and therefore prone to decay. Except metal matrices and glass masters, audio-visual carriers have an even shorter life expectancy than paper-based documents, and are especially endangered by chemical and physical decay, wrong handling and by use of poorly maintained equipment for signal retrieval.

Covering the subject of format obsolescence in chapter 4, TC03 describes audio and video recordings as machine readable documents. Even documents in perfect condition would be useless without replay machines, which makes audiovisual carriers highly endangered by format obsolescence. In practice, all analogue formats and a big part of early digital formats is already obsolete, which means that equipment is not produced any longer and support of the manufacturers is poorly available or lies in the hands of few worldwide specialists only. Therefore the solution recommended by IASA is transfer of analogue and digital materials into a true file format.

The strategy is outlined in chapter 5, by safeguarding the information by preservation of the carrier and/or by subsequent copying of the information. The life of most audio carriers cannot be extended indefinitely, availability of hardware and equipment is limited. Long-term preservation of information can only be achieved by subsequent, lossless copying from one information storage carrier/system to the next. Lossless copying can only be achieved in the digital domain, so analogue contents have to be digitised first. This principle has been generally accepted for audio preservation since ~1990.

Analogue and digital contents must be extracted from originals, analogue converted to digital, and both to file formats. This transfer is time consuming and expensive, and unlikely to be done again. Consequently original signals must be extracted and transferred in the best possible quality.

Chapter 6, selection of best copy and carrier restoration, covers the question which item to use for transfer. This is especially important if copies of various generations exist and if cooperation to other archives is required concerning this subject. In case of carrier cleaning and restoration procedures utmost care has to be taken to balance improvement of signal retrieval against possible further deterioration of the carrier. The use and wear of the original has must be kept to the minimum with all actions undertaken.

Chapter 7 covers the basic principle of optimal signal retrieval from original carriers. The digitisation process determines signal quality for the rest of document's life. The transfer may be a once-and-only process because of carrier degradation and financial constraints. It is recommended to use latest generation equipment which has to be adapted to historical formats if necessary. After transfer the originals have to be kept anyhow for later consultation.

Chapter 8 concentrates on the transfer to a new target format. It is essential, that the transfer into the digital domain is done in a completely unmodified and straight way, according to the specifications of the medium at time of its creation. In practice this means that no restoration or aesthetical improvement is permitted in an archival transfer, and that all procedures of signal processing have to be done on a second copy after transfer.

Chapter 9 deals with possible improvements in transfer technologies, which is another reason for keeping the originals as long as possible.

Chapter 10 discusses digital target formats and resolution. It is reasonable to employ openly defined formats only and to use file formats instead of data streams (e.g. CD audio stream). The use of .wav or preferably BWF is recommended.

Data reduction is covered in chapter 11. Described as a powerful tool for dissemination, it is strictly not permitted for archiving of analogue or linear digital originals, as data is omitted irretrievably. The result is quality loss and data reduction makes sources become worthless for certain analysing purposes.

Chapter 12 gives a brief introduction to digital archiving principles: Data integrity check is necessary after production and in regular intervals, and data refreshment has to be performed before content becomes irretrievable. The migration cycle to new storage systems before old systems become obsolete is discussed.

The idea of digital mass storage systems (DMSS) as storage environments is given in chapter 13, followed by chapter 14, outlining solutions before DMSSs become affordable: Small scale manual approaches to digital storage. Both chapters do not really deal with the subject matter but refer to the principle as useful in the archival strategy. Solutions based on optical storage media, such as CD-R's and DVD-R's are not considered reliable, despite in case of extensive media testing. Such procedures in term need significant investment in dedicated equipment and are time consuming.

As a matter of fact, the A/D transfer produces preservation metadata which are discussed in chapter 15. These contain details about

- the original carrier, its format and state of preservation,
- replay equipment used for transfer and its parameters
- the digital resolution, file format information and all equipment used
- the operators involved in the process
- details of the secondary information sources

Additionally a checksum should be used as a digital signature that permits authentication of the file, especially when transferred within networks.

Chapter 16 helps to develop a strategy for long term preservation. Apart from carrier degradation, recent developments suggest that format obsolescence and the associated unavailability of replay equipment may become an equal, if not greater threat for the future preservation of information. The time window for reformatting may be 15 – 20 years only.

The Publication closes with encouraging the co-operation of archives in preservation work (chapter 17) and proposes the maintaining the knowledge base of archives. *"The archive must, as a result, keep itself and its employees updated with the last scientific and technical information concerning the extraction of both primary and secondary information from carriers and improvements in preservation and restoration practices." [3].*

IASA TC03 is available as print version and as web version, and has been translated in March 2007 to German, French, Spanish, and Swedish. Translations to Russian and Chinese are available from IASA website soon.

3. IASA TC04: Guidelines on the Production and Preservation of Digital Audio Objects

IASA-TC 04 *"Guidelines on the Production and Preservation of Digital Audio Objects"* is edited by Kevin Bradley. Contributors are members of the IASA Technical Committee, as Kevin Bradley, George Brock-Nannestad, Mathew Davis, Lars Gaustad, Ian Gilmour, Michael Risnyovszky, Albrecht Häfner, Dietrich Schüller, Lloyd Stickells and Jim Wheeler, and the publication is reviewed by the IASA Technical Committee. The first version has been published in 2004.

IASA TC04 is the practical complement to IASA-TC 03. The publication is intended to provide guidance to audiovisual archivists on a professional approach to the production and preservation of digital audio objects. It is the practical outcome of the previous IASA Technical Committee paper, IASATC 03. The Guidelines address the production of digital copies from analogue originals for the purposes of preservation, the transfer of digital originals to storage systems, as well as the recording of original material in digital form intended for long-term archival storage.

After introducing the topic the guidelines start with a definition of Key Digital Principles and Standards in chapter 2. Herein the use of high quality stand alone A/D converters is recommended and the minimum specifications are defined. The file formats recommended are – being consistent with TC03 - linear PCM (.wav or preferably BWF), with a minimum resolution of 48 kHz 24 bit. No data reduction ("compression") for analogue or linear digital originals must be used.

Chapter 3 briefly discusses metadata, a shortcoming that will be extended to the actual situation in the first revision of the document, which is currently ongoing.

Chapter 4 recommends the use of Unique and Persistent Identifiers and outlines an adequate strategy.

Chapter 5 is a most useful source of practical knowledge, and information on standards and advice. It deals with signal extraction from originals and covers historical analogue as well as more modern digital formats, like microgroove discs, analogue magnetic tapes, digital magnetic carriers and optical disk media. For all these media a sequence of action is proposed, starting with the selection of best copy, cleaning, restoration, removal of storage related artefacts, choice of adequate replay equipment, speed, replay equalisation, correction for misaligned recording equipment. An estimation of time factor (the relation document's duration versus processing time for one operator) is given for each of these media. The time factor is one of the most underrated element in transfer projects, as inhomogeneous collections and historical materials might take much longer time (up to factor >3 – open ended) for digitisation and accurate documentation. A automated "factory" transfer is very cost intensive and mostly not applicable to average heritage/memory institutions.

Chapter 6 covers the digital strategy, starting with preservation target formats and systems and discussing data and audio specific storage technology. The concept of digital mass storage systems (DMSS) for archival storage is outlined, and data tape types and formats are discussed, as well as hard disk drives and arrays. A small scale manual approach to digital storage is briefly outlined, and the disadvantages of optical disks like recordable CD/DVD's are discussed, as well as the possible advantages and disadvantages of magneto-optical disks. The guidelines end with a bibliography and an index.

The publication is currently under revision, and will provide more detailed information especially concerning the chapters about Metadata (discussing Unique and Persistent Identifiers, and providing guidance on naming and numbering of files and digital works). Chapter 6 Preservation Target Formats and Systems will be structured around the functional categories identified in the Reference Model for an Open Archival Information System (OAIS, ISO 14721:2003) http://public.ccsds.org/publications/archive/650x0b1.pdf

A new chapter about, partnerships, project planning and resources will be added and will provide advice on the issues to consider if a collection manager decides to outsource all or part of the processes involved in the preservation of the audio collections.

The discussion of small scale approaches to digital storage systems will give an wider overview how to build a low cost digital management system which, while limited in scope, still adheres to the principles and quality measures identified within the publication. IASA-TC 04 is available in Spanish (2006) and Italian (2007) from http://www.iasa-web.org/technical.asp.

4. Practical Application of IASA guidelines in small scale digitisation projects

Over the past 50 years research institutes in Eastern European countries have accumulated significant collections of linguisitic, ethnomusicological and folkloristic audio material which will only survive if transferred into the digital domain in the mid-term. The Phonogrammarchiv is involved in several such digitisation projects, which have partly been funded from outside. The collections supported in strategic planning and practical implementation of digitisation include the Institutul de Etnografie si Folklor "Constantin Brăiloiu" of the Rumanian Academy of Sciences, Bucharest (Rumania), the Institute for Folk Culture Albanian Academy of Sciences, Tirana (Albania) and the Phonogrammarchiv St. Petersburg, Pushkinsky Dom, (Russia). Common problems of these institutions are factors like a lack of analogue and digital equipment, as well as missing expertise and the financial means to keeping digital data alive.

The assignment of tasks covers the long-term preservation and accessibility by digitisation of several thousand sound documents which are endangered by carrier deterioration, bad storage conditions and format obsolescence. These recordings are of unique contents, containing mostly archive-own (field-) recordings, which are incorporated in a research environment and therefore need individual transfer & scientific documentation. Basis requirements that make such a project financially rewarding are:

- a minimum size of the collection
- increase of the collection to expect
- not too many different formats to cover

If these requirements are given, an individual in-house transfer is cost-effective. A point of intersection is given at a number of about 2000 hours of audio material to be digitised, provided that educated (scientific and/or technical) staff with knowledge of the collection is available, and cooperation with local IT specialists is possible.

4.1 Assessment of the collection

The first step in starting such a project is an assessment of the collection with the aim to gain as much information as possible about the overall preservation status, the size in terms of playing time and technical parameters required for calculation of replay equipment. It is a matter of fact that the more information about the collection is available, the better the calculation of needs can be carried out.

IASA offers a special publication to examine the issues underlying the process of setting priorities for the digital transfer of analogue and digital audio content, and to deliver a statement of principles for use by sound archives in their planning for digitisation. This is the *"Task Force to establish Selection Criteria of Analogue and Digital Audio Contents for Transfer to Data Formats for Preservation Purposes"*

Helpful tools for such an assessment have been developed within the project PrestoSpace. The preservation calculator is available from http://prestospace-sam.ssl.co.uk/hosted/d14.2/newcalc.php.

Another helpful tool especially for estimation the overall preservation status of individual collections has been developed by the Indiana University Digital Library Program within the Project Sound Directions. The FACET (Field Audio Collection Evaluation Tool) can be downloaded from the Sound Directions website http://www.dlib.indiana.edu/projects/sounddirections/facet/index.shtml. This institution furthermore provides a useful reader for best practices on audiovisual archiving [4].

Additionally it is useful to make an assessment of the equipment that has been used for recording the original tapes. Although in many cases it is not possible to get information about all machines, it is still helpful to find out details about track formats and speeds that can be found within a collection. This helps to avoid miscalculation concerning playing time and equipment needs.

An assessment of the existing metadata structure is also useful and will help in calculating costs for database implementation.

4.2 Developing a Preservation Plan

In the next step a preservation plan has to be developed, proposing a prioritised sequence of actions, based on different urgencies for different parts of the collection. In such a preservation plan the focus is set on the most endangered medium with the highest scientific value and the most frequent access, balancing these factors carefully. The preservation plan should include calculations for optimising storage conditions, transferring to digital, the definition of equipment needed and finally should represent a profound basis for designing a business plan of investment.

A setup of infrastructure should include calculations for

- Analogue replay equipment
- Maintenance equipment
- Digitisation workstations
- Access stations
- Server
- Database

The preservation plan developed in our projects is based on a concept for a small scale approach to digital storage.

The simplest concept is a single operator input station with a RAID array attached. The contents of the RAID have to be regularly and at least double copied to data tapes (LTO3). As disk storage has become constantly cheaper during the last years, this is a comparatively easy and applicable solution. This setup requires a well structured plan for digitisation, as well as careful management of copy location information and version information, which is done semi-manually.

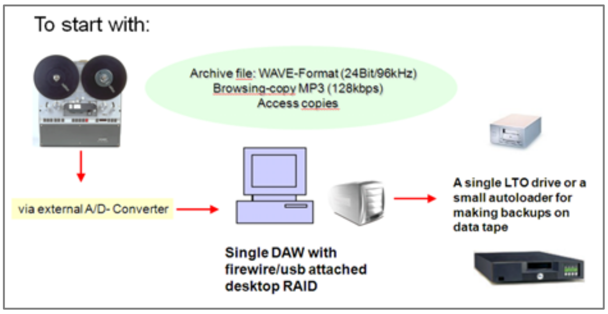


Fig. 1: Simple concept for a small scale approach to digital storage.

The system can be expanded to a small scale network for two ingest stations and one or more users, on basis of a Network Attached Storage (NAS) system with a capacity between ~ 0.5 to 20 terabyte (TB) of disk storage. In combination with an LTO autoloader this is already a midrange solution but certainly needs a higher level of administration to work properly.

To have the administration and support of such a concept guaranteed support is managed by local IT specialists, coming from the local Academies IT departments.

Whenever such manually handled solutions come into consideration, a stringent copy and safety strategy has to be implemented. This can be reached by using Unique Identifiers that can be written to the data tapes header and can be useful for data verification.

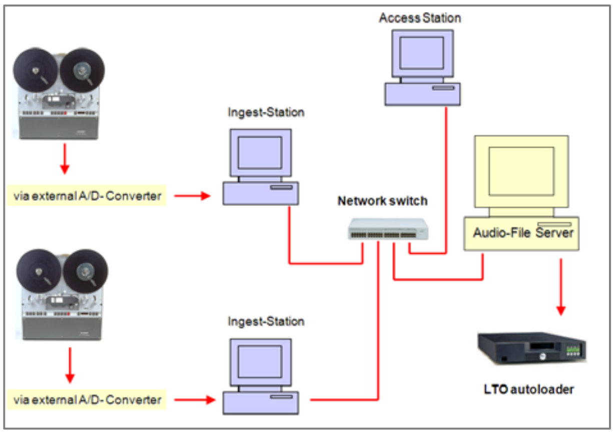


Fig. 2. Small scale network for two ingest stations and one or more users

4.3. Metadata

As dedicated tools for capturing metadata were missing in one case of our digitisation projects, we had to find a cost efficient and easily implementable answer to this important point. The focus was set on a solution that can be easily locally managed and hosted, can be easily integrated to an existing intranet, can be opened to the Internet if necessary, and provides very good safety mechanisms and data security strategies.

In cooperation with consultants we implemented a database based on widely used open source software and kept very simple and easygoing, so that it can be handled by untrained staff. The system is based on widely used open source software (Linux Ubuntu 7.04, Databasesystem MySQL, Apache2 Server, user interface based on php5).

4.4. Training

Within the projects, a 2 weeks hands-on training of technically interested/ educated staff (with practical archival experience) in the Phonogrammarchiv Vienna was calculated. The main focus of these trainings is set on the unmodified high quality transfer (conform with IASA TC03/TC04), digitisation workflow, maintenance of analogue tape machines, and documentation, especially of transfer metadata.

The digitisation process according to IASA TC04 for analogue magnetic tape is outlined in detail on the following website http://www.jazzpoparkisto.net/audio/

4.5. Practical Implementation

The practical implementation of the projects includes acquisition of adequate replay equipment. In practice, as new analogue magnetic tape machines meeting the IASA specifications are not available from the market anymore, used replay equipment has been purchased and revised to fit the necessary specifications and parameters outlined in JASATC04. The equipment has been shipped and thereafter was installed locally.

Digital equipment has been purchased from local providers, as local support is very important. After successful on-site installation the local staff was trained and a digitisation of some first analogue tapes has been carried out. Within this process the workflow for digitisation has been optimized and adapted to the needs of the archives specifically. After successful processing of some critical tapes, the first results have been presented to the department heads.

All of the digitisation projects outlined are successfully working and have already successfully digitized most of their holdings.

4.6. Subsequent Technical and Conceptual Support

The projects are subsequently supported concerning A/D transfer and technical problems on the playback side to guarantee an individual high quality transfer with optimum signal retrieval from original tapes.

Long-term service of storage facilities is solved by cooperation with local IT specialists. It is important to outline the calculation of running costs to keep digital data alive, a subject matter that will be further discussed and outlined in the revised version 2 of IASA TC04.

5. Conclusion

The standards related to the long term preservation and digitisation of sound recordings published by the Technical Committee of the International Association of Sound and Audiovisual Archives (IASA) import most powerful help and guidelines in all archival matters, be it political and strategic decisions, preservation planning and funding or the practical daily archival work. The example of practical implementation outlined above shows, that the guidelines are most useful in daily archival work and therefore represent the current standard of audiovisual archiving.

References

[1] International Association of Sound and Audiovisual Archives, Technical Committee: Standards, Recommended Practices and Strategies. *"IASA-TC 03: The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy."* Version 3, December 2005, p.3.

[2] Godsill, S.; Rayner, P.: Digital Audio Restoration - *A Statistical Model-Based Approach*. Springer Verlag, Berlin 1998, ISBN 3- 540 -76222 -1.

[3] International Association of Sound and Audiovisual Archives, Technical Committee: Standards, Recommended Practices and Strategies. *"IASA-TC 03: The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy."* Version 3, December 2005, p.12.

[4] Casey, Mike; Gordon, Bruce, Editors: *"Sound Directions. Best Practices for Audio Preservation."* Indiana University, 2007.

Semi-automatic metadata extraction from shellac and vinyl disc

Nicola Orio
Information Management Systems Research Group
Dept. of Information Engineering
University of Padova

Lauro Snidaro
Laboratorio AVIRES
Dept. of Mathematics nd Informatics
University of Udine

Sergio Canazza
Laboratorio AVIRES
Dept. of Historical and Documentary Science
University of Udine

Abstract— During the process of active preservation, the original analogue document – multimedia in itself, because it is made up of the audio signal, static images (label, case, carrier corruptions, etc.), text (attachments), smell (mould, etc.) – is converted into a digital *unimedia* document. This *projection* of a multidimensional object (the original document) into a one-dimensional space (the bit flow) produces a large and various set of digital documents, which are made up of the audio signal, the metadata and the contextual information. In medium/large archives it is unrealistic to manually extract the metadata from video shootings and photos. The goal of this work is to presents an informatics system able to extract in a semi-automatic way metadata from photos and video shootings of phonographic discs.

***Index Terms*—New technologies for audio heritage, Semi-automatic metadata annotation, Audio alignment, Computer vision.**

I. INTRODUCTION

Since the paper used in 1860 (first audio *recording*[1] by Édouard-Léon Scott de Martinville *Au Clair de la Lune* using his phonautograph) to the modern Blu-ray Disc, what we have in the audio carriers field today is a Tower of Babel: a bunch of incompatible analog and digital approaches (paper, wire, wax cylinder, shellac disc, film, magnetic tape, vinyl record, magnetic and optical disc, etc., etc.) without standard players able to read all of them. The wide time span in which these formats have been developed makes it even harder to select the correct playing format for each carrier. It should be clear the importance of transfer into the digital domain (active preservation), namely for in carriers in risk of disappearing, respecting the indications of the international archive community (see, at least, [1], [2], [3], [4], [5]).

The opening up of archives and libraries to a large telecoms community, which has been made available through their integration into the Internet, represents a fundamental impulse for cultural and didactic development. Guaranteeing an easy and ample dissemination of some of the fundamental moments of the musical culture of our times is an act of democracy which cannot be renounced and which must be assured to future generations, even through the creation of new instruments for the acquisition, preservation and transmission of information. This is a crucial point, which is nowadays the core of reflection of the international archive community. If, on one hand, scholars and the general public have begun paying greater attention to the recordings of artistic events, on the other, the systematic preservation and access to these documents is complicated by their diversified nature and amount.

It is well-known that the recording of an event can never be a neutral operation, since the timbre quality and the plastic value of the recorded sound, which are of great importance in contemporary music (electro-acoustic, pop/rock, ethno music) are already determined by the choice of the number and arrangement of the microphones used during the recording. In particular, in cases of a non traditional stationing of the orchestra players or of pieces based on improvisation, the positioning of the microphone

1 Unlike Edison's similar 1877 invention, the phonograph, the phonautograph only created visual images of the sound and did not have the ability to play back its recordings. Scott de Martinville's device was used only for scientific investigations of sound waves.

according to purely documental and presumably *neutral* criteria can be a naïve solution, which in practice sets serious limits to the identification of the piece. Moreover, the audio processing carried out by the *tonmeister* is a real interpretative element added to the recording of the event. Thus, musicological and historic-critical competence becomes essential for the individuation and correct cataloguing of the information contained in audio documents. The commingling of a technical and scientific formation with historic-philological knowledge also becomes essential for preservative re-recording operations, which do not coincide completely with pure A/D transfer, as it is, unfortunately, often thought.

The increased dimensionality of the data contained within an audio digital library should be dealt with by means of automatic annotation. The auditory information contained in the audio medium can be augmented with cross-modal cues. For instance, the visual and textual information carried by the cover, the label and possible attachments has to be acquired through photos and/or videos. The storage and representation of this valuable information is common practice and is usually based on well-known techniques for image and video processing, such as OCR, video segmentation and so on. We believe that it is interesting as well, even if not studied yet, to deal with other visual information regarding the carrier corruption and imperfection occurred during the A/D conversion.
This work presents a set of tools able to extract, in a semi-automatic way, metadata from photos and video shootings of audio carriers (Section II). Moreover, we introduce a system for reconstructing the audio signal from the photo of the disc surface (Section III). Section IV describes alignment techniques useful in the comparison of alternative digital acquisitions. Finally, Section V provides a case study in which an alignment tool is used in order to notate disc corruptions.

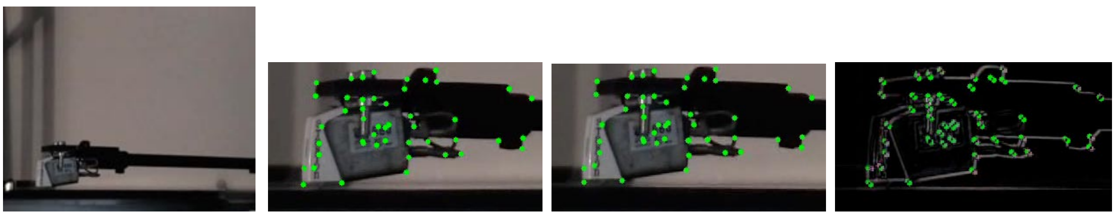
II. METADATA EXTRACTION FROM CARRIER VIDEOSHOOTINGS AND PHOTOS

Computer vision algorithms and techniques can be applied for the automatic extraction of relevant metadata to be associated to the auditory information.

A. Warped discs

The characteristics of the arm oscillations can be related to pitch variation of the audio signal. Therefore, they constitute valuable metadata for audio signal restoration processes. Thus we here propose computer vision techniques for the automatic analysis and annotation of videos of rotating discs.

We have employed a feature tracking algorithm known as the Lucas-Kanade tracker [14]. The algorithm locates feature points on the image to be tracked between consecutive frames. The technique, initially conceived for image registration, is here employed as a feature tracker to keep track of the position of the features from a frame to the following one.
Figure 1 shows some frames from one of the sequences used in the experiments: (b) shows the lowest position of the arm's head in one oscillation and (c) the highest position, where the Lucas-Kanade features can be seen on the arm's head while being tracked through the oscillation. Even if from the Figure 1 the differences between the highest and lowest positions are almost noticeable (see the differences between them in (d)), nevertheless our approach is able to track them clearly, as shown in Figure 2.



(a) (b) (c) (d)

Figure 1. Processed frames from a video of a oscillating record player's arm. (a) Photo of the turntable arm; (b) Lowest position of the arm in a oscillation, (c) its highest position. (b) and (c) show Lucas-Kanade features detected on the arms' head and tracked through the oscillation. (d) shows the differences between lowest and highest positions.

In the experiments, the Lucas-Kanade tracker has correctly tracked the features detected in the first frame of the video sequences. The tracker has thus allowed to analyze the temporal evolution of the position of the features on the arm's head while the record player was playing a severely deformed disc.

Figure 2 shows the temporal evolution of the y coordinate of a feature located on the arm's head. The x-axis shows the number of frames and the y-axis reports the position in pixels on the image plane. The oscillatory evolution is clearly visible. There is a 29 frames gap between (a) and (d), and this is consistent with the distance between the highest peaks and the lowest peaks in Figure 2.

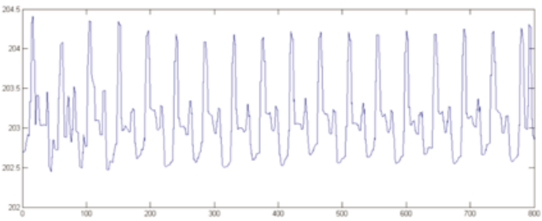


Figure 2. Temporal evolution of the y coordinate of a Lucas-Kanade feature located on the arm's head. It can be seen clearly how the oscillations indicate a deformed disc.

B. *Off-centered discs*

Interesting properties of phonograph records can be automatically extracted by analyzing a picture of it. For example, we wanted to calculate the eccentricity of the disc, that is, the offset between the spindle hole axis and the exact central rotation axis. This production flaw, which could affect individual copies or entire stocks of records, is responsible for the well-know warp effect that introduces a pitch variation in the audio signal.

To accomplish this automatically we have exploited the consolidated literature in iris detection, which is a required processing step for each iris recognition system [7].

Since our problem shares the same lucky circular properties of the problem of iris detection, we have employed the integrodifferential operator which was developed for detecting the pupillary boundary and the outer boundary of the iris [7].

The integrodifferential operator has the following form:

$$\max_{(r,x_0,y_0)} \left| G_\sigma(r) * \frac{\partial}{\partial r} \oint_{r,x_0,y_0} \frac{I(x,y)}{2\pi r} ds \right| \qquad (1)$$

The operator is computed over the image *I(x,y)* where it searches for the maximum of the blurred partial derivative, with respect to the radius *r*, of the normalized circular integral of radius *r* and center coordinates (x_0, y_0) calculated on *I(x,y)*. The blur is obtained through convolution with a Gaussian smoothing function of scale σ.

In other words, the operator works as circular edge detector and it provides the center coordinates and the radius of the strongest circular edge in the image. In our implementation, we extracted the outer contour of the disc first and then rerun the operator on the image for detecting the spindle hole contour as shown in Figure 3. The second pass can be computed very fast as it takes advantage of the known geometrical properties of vinyl discs. That is, once the outer boundary has been detected the spindle hole contour can be searched in a subregion of the image inside the outer contour.

In our set-up, the disc was laying on a plane parallel to the image and the spindle hole was on-axis with the camera's optical axis. Although this constraint isn't too much restrictive for a dedicated set-up in an audio laboratory, a step further can be taken by removing this assumption and considering perspective deformations given by out-of-axis images as discussed in [8], [9].

Figure 3. Disc and spindle hole contours automatically detected via the integrodifferential operator.

Having detected the outer boundary of the disc and the spindle hole contour, the calculation of the offset between their centres is trivial. In the experiment reported in Figure 3, the estimated offset was 1,414 pixels corresponding to 0.22 cm.

III. *PHOTOS OF GHOSTS* (PHOTOS OF GROOVES AND HOLES, SUPPORTING TRACKS SEPARATION)

Nowadays, automatic text scanning and optical character recognition are in wide use at major libraries: unlike texts, A/D transfer of historical sound recordings is often an invasive process.
As it is well-known, several phonographs (*laser turntable*) exist able to play gramophone records using a laser beam as the pickup. This playback system has the advantage of never physically touching the record during playback: the laser beam traces the signal undulations in the record, without friction. Unfortunately, the laser turntables are constrained to the reflected laser spot only and are susceptible to damage and debris and very sensitive to surface reflectivity.
Digital image processing techniques can be applied to the problem of extracting audio data from recorded grooves, acquired using an electronic camera or other imaging system. The images can be processed to extract the audio data. Such an approach offers a way to provide non-contact reconstruction and may in principle sample any region of the groove, also in the case of a broken disc. These scanning methods may form the basis of a strategy for: a) larger scale A/D transfer of mechanical recordings which retains maximal information (2D or 3D model of the grooves) about the native carrier; b) the active preservation of carriers with heavy degradation (breakage, flaking, exudation).
In literature there are several approaches to this problem (see, at least, [10], [11], [12] and [13]). The authors have developed an HW/SW system (*Photos of GHOSTS*, [14]): a) able, automatically, to recognize different rpm and to perform tracks separation; it doesn't require human intervention; c) works with low-cost hardware; d) is robust with respect to dust and scratches; e) outputs de-noised and de-wowed audio, by means of novel restoration algorithms. An equalization curve choice by the user is possible: the system has hundreds of curves stored, each one with appropriated references (date, company, roll-off, turnover). Moreover, *Photos of GHOSTS* allows the user to process the signal by means of several audio restoration algorithms [15], [16].

The system uses a customized scanner device with rotating lamp carriage in order to position every sector with the optimal alignment relative to the lamp (coaxially incident light). The software automatically finds the record center and radius from the scanned data, for performing the groove rectification and for separating the tracks. Starting from the light intensity curve of the pixels in the scanned image, the groove is modeled and, so, the audio samples are obtained (Figure 4).

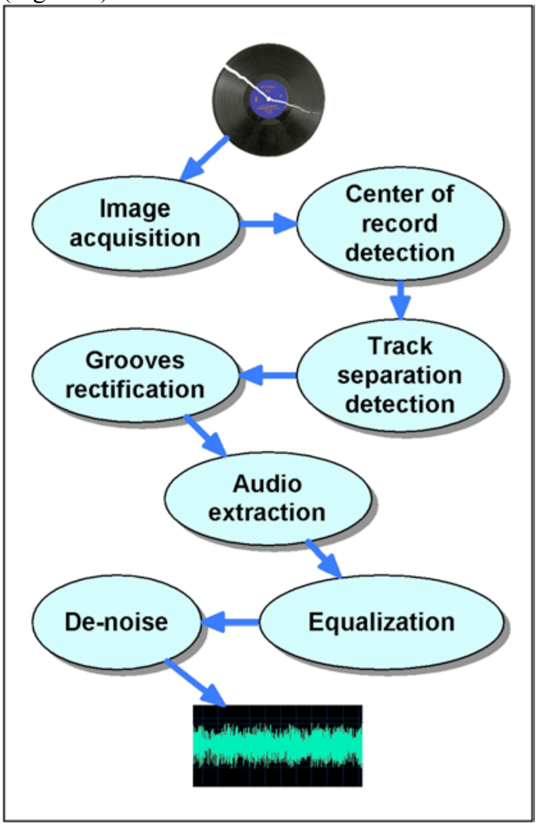


Figure 4. Photos of GHOSTS schema.

IV. AUDIO ALIGNMENT

The typical application of audio alignment is the comparison of two alternative performances of the same music work. This comparison can be helpful to musicologists for studying the style of different conductors and performers, and it can also be exploited to re-synthesize performances adding new expressive parameters. In the case of classical music, alignment can be carried out also between the recording of the performance and a digital representation of the score, yet audio to audio alignment may be the only option for genres that are not commonly represented by a standard notation, such as ethnic or electro-acoustic music. The alignment of two audio recordings can be a useful tool also when two different versions of the same recording session are to be compared. For instance, in the case of electro-acoustic music, the available recordings of a given work may differ because of different post processing and editing that have been applied before publication [18]. In this case, alignment allows musicologists to highlight possible cuts and insertions of new material in the recordings, to detect the usage of previous released material inside a new composition, and to compare the temporal and spectral features in corresponding parts also when they have different playback speed.

We propose to apply alignment techniques to the comparison of alternative digital acquisitions of the same disc. In particular, the technique based on digital images which is presented in this paper is compared to the acquisition based on analogue playback. It is likely that the recording speeds differ slightly depending on the technique and that there can be local differences depending on the quality of the analog equipment. Moreover, the two approaches may give different results in terms of robustness to local damages on the record surface. For this reason, we propose to use automatic alignment as a tool to compare the characteristics of digital acquisition of a given record and to evaluate objectively the quality of the proposed technique.

Audio to audio matching is usually based on a preprocessing of the recordings in order to extract relevant features that are able to generalize their main characteristics. A popular descriptor is the chroma-based representations, here the basic idea is that all the components of the spectrum are conflated into a single octave, obtaining a particular signature of a polyphonic signal. Alternatively, as presented in [19], audio recordings can be segmented in coherent parts with stable pitch components, and a set of bandpass filters are computed for each segment around the main peaks in the frequency domain. Once a set of descriptors is computed from the two audio signals, the global matching can be carried out using dynamic programming approaches to compute the local and global distance between frames in the recording, for instance Dynamic Time Warping (DTW), or statistical modeling of the temporal and spectral differences between the two recordings, for which the most used tools are Hidden Markov Models (HMMs). Both approaches are quite popular in the speech recognition research area [20]. For example, a variant of DTW has been proposed in [21] for off-line alignment using chroma features, while a real time version of DTW has been presented in [22]. An approach to alignment based on HMMs is described in [19].

Both in the case of DTW and HMMs, the global alignment is computed from a local distance using a dynamic programming approach. The main difference is that HMMs require that a model is built from one of the recordings, which becomes the reference signal against which the other recording is compared, while DTW can be carried out directly from the signal parameters without the need of using a particular recording as the reference. Another important difference is that HMMs need to be trained with a large number of examples, which might not be available in some application domains, while DTW is simply based on the notion of local distance between audio frames of the two recordings. For these reasons, DTW has been used in our work to compute the alignment. Another reason is that, for the musicologists who analyze the results of the automatic alignment, it is more intuitive to think about distances rather than marginal probabilities.

The first step in the definition of a distance between two recordings regards the choice of the acoustic parameters that are to be used. Given the relevance of spectral information, the similarity function is normally based on the frequency representation of the signal. In order to highlight also short local mismatches due to small scratches on the record surface, we choose to use small windows of the signal, of 2048 points with a sampling rate of 44.1 kHz, using an hopsize between two subsequent windows of 1024 points. These parameters give a time resolution of the alignment of about 23 milliseconds.

After choosing how to describe the digital recordings, a suitable distance function has to be chosen. Many distances have been proposed in the literature to measure the distance between two spectra, ranging from cross correlation, spectral flux, to L1 and L2 norms. We propose to use the cosine of the angle between the vectors representing the amplitude of the Fourier transform, which is a well known measure used typically in information retrieval. Thus, given two recordings f and g, the local distance $d(m,n)$ between two frames can be computed according to equation

$$d(m,n) = \frac{\sum_{i=1}^{K} F_m(i)\, G_n(i)}{\|F_m\|\ \|G_n\|} \tag{2}$$

where F_m (G_n) is the magnitude spectrum of frame m (n) of recording f (g), while in our application K = 2048 points. Local distance can be represented by a distance matrix, as shown in Figure 5, which can be used as a visual representation of the similarities between two recordings. As it can be seen from the Figure 5, the main similarities are along the main diagonal, where large dark squares correspond to long sustained notes and brighter areas represent a low degree of similarity between two frames. In practice, the local distance needs to be computed only in proximity of the main diagonal, in order to reduce computational cost.

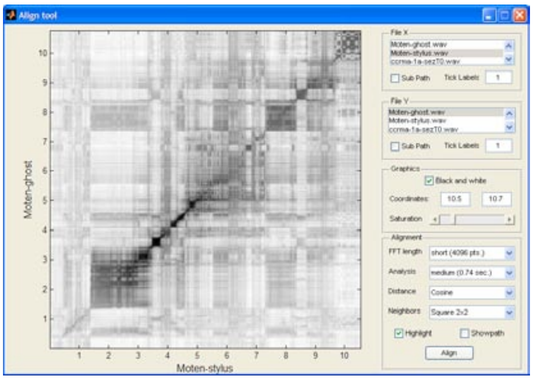


Figure 5. Visual representation of the similarities between two audio signals. X-axis: the audio signal generated from a photo of the disc by means Photos of GHOSTS system (see Section III); y-axis: audio signal extracted by means of turntable.

After the local distance matrix is computed, DTW finds the best aligning path according to equations

$$c(m,n) = \min \begin{cases} c(m-1,n-1) + 1.5\ d(m,n) \\ c(m-1,n) + d(m,n) \\ c(m,n-1) + d(m,n) \end{cases} \tag{3}$$

$$p(m,n) = \arg\min \begin{cases} c(m-1,n-1) + 1.5\ d(m,n) \\ c(m-1,n) + d(m,n) \\ c(m,n-1) + d(m,n) \end{cases} \tag{4}$$

where *c(m,n)* is the cumulative distance between the two recordings, computed for each couple of frames. It is possible to compute the global optimal path that starts in point [1,1] and stops in any chosen point through a backtracking procedure that exploits the information stored in *p*(*m,n*). It has to be noted that there have been proposed many different combinations of neighbor points to compute the minimization. The results presented in this paper have been computed using this equation, which is based on just three neighbors located on a square.

V. CASE STUDY

As case study, we selected the double-sided 78 rpm shellac disc Okeh 8457 – OK 8102 and put our attention on the song *A Chattanooga Blues*.
The performers are Mary H. Bradford (v) with Bennie Moten's Kansas City Orchestra: Lammar Wright, c; Thamon Hayes, tb-1; Woodie Walder, cl-1; Bennie Moten, p/ldr; George Tall, bj-1; Willie Hall, d-1. September 1923. This is an acoustic recordings (made prior to the use of microphones). Bennie Moten is today remembered as the leader of a band that partly became the nucleus of the original Count Basie Orchestra. He was a fine ragtime-oriented pianist who led the top territory band of the 1920s, an orchestra that really set the standard for Kansas City jazz. Moten formed his group (originally a sextet) in 1922 and the following year they made their first recordings.
The audio signal was extracted in two ways:
1) by means of the Rek-O-Kut-Rodine 3 turntable; the A/D transfer was carried out with RME Fireface 400 at 44.1kHz/16bit. We didn't applied any equalization curve (acoustic recording).
2) using *Photos of GHOSTS* system (the photo was at 4800 dpi, 8 bit grayscale, without digital correction).
Finally, the alignment method presented in Section IV was used in order to compare the differences/similarities between these two audio signals. In this way, interesting metadata (about the A/D transfer process and the original carrier) can be extracted.
Alignment curve: by comparing the two signals, it is possible to point out the discrepancies between the angular velocities used during the disc playing (Figure 6). The virtual *velocity* of the *Photos of GHOSTS* is perfectly constant, of course (given by the number of pixels/second read by the software); therefore, the blue curve

shows the imperfections of the A/D transfer system (acceleration/deceleration of the turntable during the playing). In our case, the velocity of the audio signal generated by *Photos of GHOSTS* is greater then that extracted with the turntable, despite we set both to 80 rpm (1923 USA Okeh acoustic recording). In this way, we have a tool for taking into account some imperfections of the A/D transfer process.

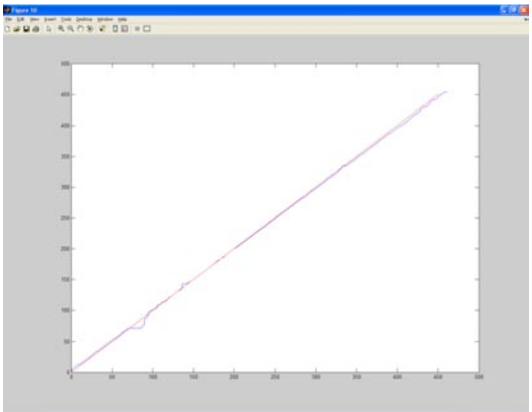


Figure 6. Alignment curve (blue curve), in comparison with the bisector (red line). X-axis: audio signal generated by photo; y-axis: audio signal extracted by means of turntable.

Visual representation of the similarities. Figure 7 shows the main similarities (dark areas): brighter areas represent a low degree of similarity between two frames. In the middle of the excerpt there are areas with a low similarity degree: in fact, in this interval the voice recorded in the signal is very distorted. These distortions are *performed* in different manners by the two systems. In this way, we have a tool able to describe serious corruptions of the recording.

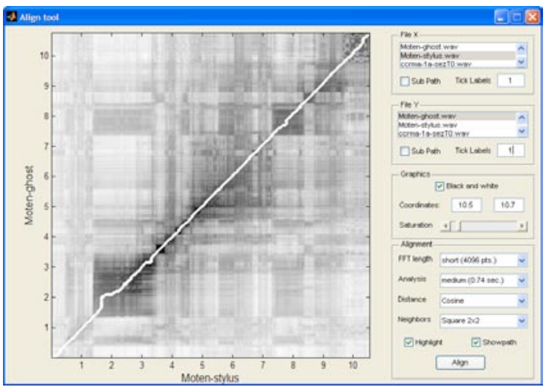


Figure 7. Visual representation of the similarities between two audio signals. X-axis: audio signal generated by photo; y-axis: audio signal extracted by means of turntable. The local distance in proximity of main diagonal is highlighted in white.

Graph of the differences. Figure 8 shows the similarities and the differences between the two signals. Because the signal generated by *Photos of GHOSTS* is very different in proximities of local disturbances (scratches and crackles). The local minimum values of the function plotted in Figure 8 give a list of the disc local corruptions.

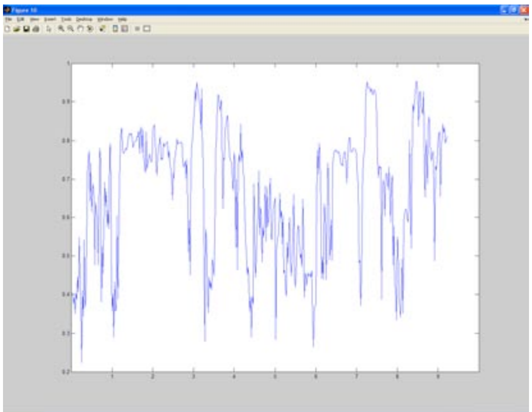


Figure 8. The graph of the differences between the two audio signals along the alignment curve. X-axis: audio signal generated by photo; y-axis: a similarity degree scaled from 0 to 1.

VI. Conclusion

In the phonographic discs archives, the information written on the edition containers (cases, envelopes, boxes), on the label and on the attachments (text and images) are usually stored in the preservation copy as static images. In this paper, we propose that this important information can be integrated with additional metadata that describe carrier corruptions and imperfections in A/D conversion. For example, the video shooting of the disc transfer – synchronized with the audio signal – ensures to preserve all the information regarding the status of carrier and its relationship with the audio quality of the digital recording.

In medium/large archives it is unrealistic to manually extract the metadata from video shootings and photos. This work presented automatic tools able to extract in semi-automatic way metadata from photos and video shootings of phonographic discs.

The processing described above can be performed on-line in real-time. In particular, the experiments have been carried out on video shooting use 320x240 resolution of video sequences with an above real-time frame rate processing performance of 50 frames/sec on a 3

GHz single processor machine. Video sequences have been acquired with a consumer handycam at PAL resolution and subsequently rescaled and compressed into DivX video files at medium-high quality setting. For the reconstruction of audio signal from photo, a customized HP ScanJet 4890 Photo is used, set with 8 bit (256 grayscale levels) and with a resolution of 4800 dpi, without digital correction.
The applications have been coded in C++ and Matlab. In addition, no particular setup was required for this experiment. The algorithms are robust to different lighting conditions. This would be a practical set-up for audio laboratories and audio digital libraries.

REFERENCES

[1] AES22-1997: *AES recommended practice for audio preservation and restoration – Storage and handling – Storage of polyester-base magnetic tape*, 2003.

[2] AES49: *AES standard for audio preservation and restoration – Magnetic tape – Care and handling practices for extended usage*, 2005.

[3] AES-11id: *AES Information document for Preservation of audio recordings – Extended term storage environment for multiple media archives*, 2006.

[4] D. Schüller, "Preserving the Facts for the Future: Principles and Practices for the Transfer of Analog Audio Documents into the Digital Domain". *Journal of the Audio Engineering Society*, vol. 49, n. 7-8, 2001, pp. 618-621.

[5] G. Boston, George (ed), *Safeguarding the Documentary Heritage. A guide to Standards, Recommended Practices and Reference Literature Related to the Preservation of Documents of all kinds*. UNESCO, Paris, 1998.

[6] J. Shi, and C. Tomasi, "Good Features to Track", IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WS, June 1994, pp. 593-600.

[7] Y. Liu, D. Zhang, G. Lu, and W.Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition*, vol. 40, n. 1, June 2003, pp. 262-282.

[8] J. Daugman, "How iris recognition works", *IEEE Transactions on circuits and systems for video technology*, vol. 14, n. 1, January 2004, pp. 21-30.

[9] J. Daugman, "New methods in iris recognition", *IEEE Transactions on system, man, and cybernetics – part B: cybernetics*, vol. 37, n. 5, 2007, pp. 1167-1175.

[10] V. Fedeyev, and C. Haber, "Reconstruction of mechanically recorded sound by image processing", *Journal of the Audio Engineering Society*, vol. 51, n. 12, 2003, pp. 1172-1185.

[11] V. Fadeyev, and C. Haber, Reconstruction of Mechanically Recorded Sound. *Lawrence Berkeley National Laboratory Technical Report 51983*, 2003.

[12] S. Cavaglieri, O. Johnsen, and F. Bapst, "Optical Retrieval and Storage of Analog Sound Recordings". *Proceedings of AES 20th International Conference*, Budapest, Hungary, October 5–7, 2001.

[13] S. Stotzer, O. Johnsen, F. Bapst, C. Sudan, and R. Ingol, "Phonographic Sound Extraction Using Image and Signal Processing", *Proceedings of ICASSP*, vol. 4, May 17-21, 2004, pp. 289-292.

[14] L. Snidaro, and G.L. Foresti, "Real-time thresholding with Euler numbers", *Pattern Recognition Letters*, vol. 24, n. 9-10, June 2003, pp. 1533-1544.

[15] S. Canazza, G. Coraddu, G. De Poli, and G.A. Mian, "Objective and subjective comparison of audio restoration methods", *Journal of New Music Research*, vol. 30, n. 1, 2001, pp. 93-102.

[16] A. Bari, S. Canazza, G. De Poli, and G.A. Mian, "Toward a Methodology for the Restoration of Electro-Acoustic Music", *Journal of New Music Research*, vol.30, n. 4, pp. 365-374.

[17] S. Canazza, G. Ferrin, and L. Snidaro, "*Photos of GHOSTS*: Photos of Grooves and HOles, Supporting Tracks Separation", *Proceedings of XVII CIM*, accepted for publication, 2008.

[18] N. Orio, and L. Zattra, "Audio matching for the philological analysis of electro-acoustic music". *Proceedings of the International Computer Music Conference*, 2007, pp. 157-164.

[19] R. Miotto, and N. Orio, "Automatic identification of music works through audio matching". *Proceedings of 11th European Conference on Digital Libraries*, 2007, pp. 124-135.

[20] L. Rabiner, and B. Juang, *Fundamentals of Speech Recognition*. Prentice-Hall, Englewood Cliffs, NJ, 1993.

[21] M. Müller, F. Kurth, and M. Clausen. "Audio matching via chroma-based statistical features", *Proceedings of the International Conference of Music Information Retrieval*, 2005, pp. 288-295.

[22] S. Dixon, and G. Widmer. MATCH: a music alignment tool chest. *Proceedings of the International Conference of Music Information Retrieval*, 2005, pp. 492-497.

CD and DVD preservation issues

Marco Righi

Workshop on Digital Preservation Weaving Factory
for Analogue Audio Collections

ISTI • Area della Ricerca CNR
via G. Moruzzi 1, 56124 PISA, Italy

***Abstract*- This paper provides basic advice for CD/DVD archives maintenance.**

I. INTRODUCTION

It's not an easy task to insure the integrity during time of the information stored within CDs and DVDs: these supports actually degrade and may become unreadable. Many studies related to disc stability over time already exist: [1][2][3][4][5][6]. The subject is interesting not only for archivists but for anyone who keeps information stored on these kinds of carries.

Unfortunately a shared opinion associates an unexceptionable reliability to that optical supports; instead we can notice, as precisely documented in the recalled references, that CDs and DVDs can become unreadable within a few years. With reference to [6] we can see that, considering a sample of 125 pre-printed Audio CD out of the Library's collection of 60,000 CDs (Library of Congress, Washington DC, USA), some of them started presenting unrecoverable errors within less than 10 years from their manufacturing. In the case of an audio CD, that does not imply the impossibility of its utilization: indeed the human ear can even not notice such errors. The reading devices in fact are built in such a way as to correct or hide errors whenever possible. In the case of unrecoverable errors, sometimes not audible by ear as well, a loss of information with respect to the original contents happens and it will affect and propagate on any subsequent copy of the support.

The situation is quite worse for CD-R's, sometimes used for creating backup copies or to store original works. Table I shows the results of some tests taken from reference [5]. The documents under examination belong to the *Archivio di Stato of Rome*. The column on the right reports the number of CD's with unrecoverable errors.

This different temporal duration is mainly due to the different materials which pre-printed CDs and CD-R are made of [22]; but it depends also on the information storage technique used: in the case on the formers it is an actual printing from a master while for the latters the storage is achieved through the modulation of the power of a laser beam and the subsequent creation of pits (tiny holes on the surface).

Even though DVDs are based on a more modern technology, some tests based on artificial aging techniques [7] highlight how even these supports are liable to decay.

TABLE I
SOME TESTS RELATED TO THE ARCHIVIO DI STATO OF ROME

Collection	Number of CD-R of the collection	Number of critical CD-R
Mappe Gregoriano	15	6
Gregoriano Mappette	3	1
Alessandrino	6	1
Pergamene	50	14
Urbano	3	1
Preziosi	21	4
Fondo Bazzani	32	5

II. INFORMATION STORAGE ON AUDIO AND NON-AUDIO CD'S

The information storage format on an audio CD differs from the one used to store other kinds of data on a non audio CD, e.g. text or programs. This happens because the mechanisms allowing the correct reading of an audio CD are not enough reliable for the storage of non audio CD related data [8][9][10].

In a CD in fact the data block is composed starting from a frame. A frame is composed of 24 bytes that correspond, in an audio CD, to 6 stereo samples. A frame represents the minimal information unit on a CD. However it can be addressed only in sets of 98 units. A group of 75 frame sets, each set composed of 98 frames, is called a sector [10].

The coding requires the addition of 8 parity bytes, 1 subcode byte, 1 byte of synchronism and 102 bits used for frame merging to the 24 bytes of data. Before the writing of the frame some of these bits, as shown in Table II and explained into the CD standard documentation [9], are modulated with the EFM technique.

We can observer that in order to store 192 bits of information we use 588 bits. The method used for error correction is named Cross Interleave Reed-Solomon Coding (CIRC). The effectiveness of this system is noteworthy: for each frame the 24 bytes of data are sent to the first Reed-Solomon decoder, which using the first 4 extra bytes is able to spot and correct an error every 32 bytes. Later the 24 bytes of data and the remaining 4 extra bytes are sent, with different timing intervals, to the second R-S decoder. The byte interleaving allows to decompose burst errors, that is on many consecutive bytes, into many errors involving a single byte for each block. In presence of errors the second decoder R-S uses the last 4 bytes to correct the 24 data bytes. After having been deinterlaced, to restore the starting order, the data can be finally sent to the output.

TABLE II
THE FRAME OF THE CD

Description	Number of bit	EFM bit
Data	192	336
Q parity	32	56
P parity	32	56
Subcode	8	14
Sync word		24
	Total	588

CD specifications allow up to 220 errors per second. The CIRC algorithm is almost always able to perfectly fix these errors and to provide correct data. Errors extended up to 450 consecutive bytes, thanks to interleaving, are fixed without any information loss. It is still possible that errors are not recognised, but usually it is the case of little "clicks" that can be barely perceived when heard by the human ear.

The higher level of control that is mandatory to ensure the correctness of the information written on non audio CD brought to use 288 additional bytes in a given sector to implement a further level of data correction. We have in fact 4 bytes of EDC (Error Detecting Code) containing the CRC of the preceding data and 276 bytes of ECC (Error Correcting Code) that, by using an interleaving technique and the Reed-Solomon codes, allow the correction of the errors that escaped the previous controls. The probability of not corrected or not detected errors is extremely low at this point.

The different CD's structures are highlighted also in Figure 1, depicting the relations between the various CD standards.

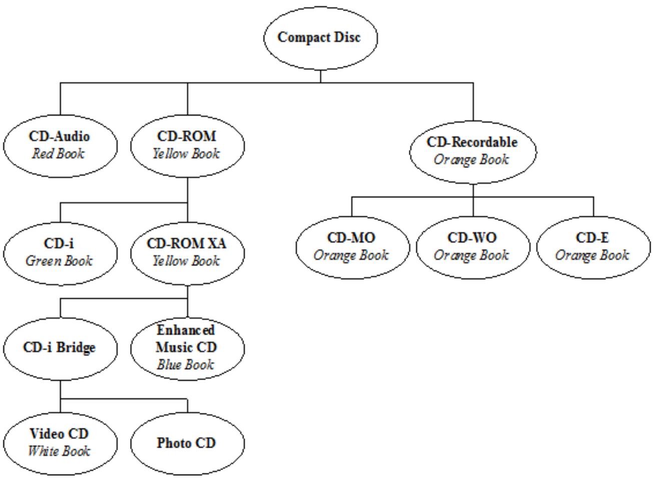


Figure 1. The standard of the CD technology.

Storage specifications on non audio CD's are also enriched with the sector addressing mechanism. On an audio CD the information addressing is entrusted to the subcode's q-bits. In order to access this information it is mandatory to read 98 frames [9], , as represented in Figure 2 [9].

A CD is characterized by a speed of 7350 frames/sec and consequently the subcode is refreshed 75 times per second.

The temporal information, used to address the bytes, is hence distributed information.

Such a situation allows the use of the CD in agreement with its original conception that was initially conceived the CD as a replacement for the vinyl disc. Later the CD was considered as a general purpose storage support which allowed storing any type of data, such as programs or pictures.

For such a reason each single non audio CD sector contains information related to its address, as depicted in Figure 3 [10].

III. DVD INFORMATION STORAGE

The DVD, described in documents [11][12][13][14][15][16][17][18][19][20][21] is characterized by a totally different structure fort he storage of information.

Standard DVD books provide specifications related to the so-called low-level, which deescribe the physical specifications. At a higher level of specification there is the description of the file system that allows the management of the stored information.

To analyze the advantages that the DVD owns with respect to the CD we must refer to the following publications:

- Book A, containing the DVD-ROM specifications
- Book B, containing the DVD-Video specifications
- Book C, containing the DVD-Audio specifications

Figure 4 [10] shows the complete compatibility between the different DVD types formats both at physical and at the file system level.

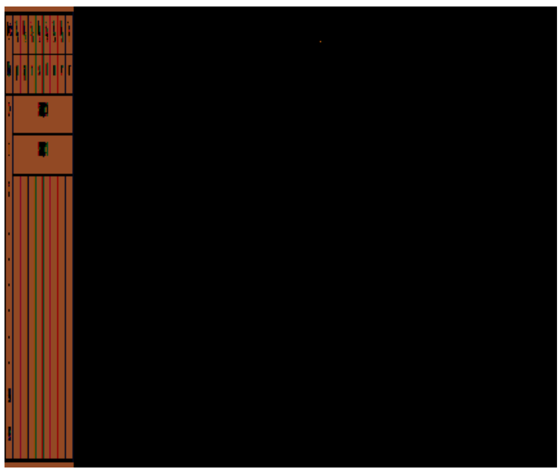


Figure 2. The subcode of the CD.

This fact allows the DVD-ROM, DVD-Audio and DVD-Video readers to access the information stored within the disc.

In the DVD a sector is composed of 2064 bytes, according to the following schema: 2048 for data, 4 bytes for error detection and 12 bytes reserved for other tasks. Each data block is represented by 16 sectors, each of them further decomposed in 12 blocks, and worked out by a scrambling algorithm.

Sector: 2352 bytes							
Sync	Header		User Data	EDC	Intermediate	P-Parity	Q-Parity
	Sector Address	Mode					
12 bytes	3 bytes	1 (01)-byte	2048 bytes	4 bytes	8 bytes	172 bytes	104 bytes
0 11	12 14	15	16 2063	2064 2067	2068 2075	2076 2247	2248 2351

Figure 3. The CD mode 1.

On the scrambling sectors the ECC codes (referred as PI and PO) are calculated. As depicted in Figure 5 [19] PI codes are written in 10 new columns and the PO codes in 16 new lines. The PI and PO codes make up the RS-PC (Reed-Solomon Product Code).

IV. CD AND DVD STATUS DIAGNOSIS PARAMETERS

In order to estimate the life expectancy of a support it would be mandatory to know its history: preservation conditions and the writing mode in the case of a CD-R or DVD±R support. Often this information is not available. To estimate the conservation state of a disc it is therefore important to check the room in which it was stored.

For example, in rooms with high humidity levels or overheated, as described in [7][23] [24], the supports have a higher probability of rapid decay. Figure 6 [22] shows an example of CD-ROT and rolling.

Figure 4. The standard of the DVD technology [10].

To check the current conditions of a CD or DVD it is possible to use specific devices that allow the analysis of some basic parameters [25]. Regarding CDs we take into consideration physical parameters (like disc eccentricity and its dimensions, pit and land size, reflectivity level of the disc and jitter signal) and logical parameters (like the BLER and the errors E11, E21, E31, E12, E22 and E32).

Figure 5. ECC Block of the DVD.

For an accurate reliability analysis of a CD-R it is also mandatory to measure some parameters before and after the burning process. Among these parameters we find the land reflectivity percentage and the ATER that must be less than 10%.

In the DVDs the physical parameters are similar to those used for the CDs, while at the logical level, because of the format diversity, the parameters are mostly different. In fact in the DVD we don't use any BLER or E32, while we rely instead on other parameters, for instance the PIE, the POE, the PISum8 (based on 8 consecutive ECC blocks). Works [1][5][7] on longevity take into consideration at least the following parameters.

For CDs:

- BLER: it is proportional to the number of data blocks containing at least one error. The BLER is defined as the error rate calculated as the sum of the E11, E21 and E31 errors per second. According to the CD specifications, the BLER should not be superior to 220 blocks per second. The maximum value of the BLER corresponds to the maximum rate of errors detected on the disc.
- E32: these are errors that the reading device cannot correct. The disc areas where these errors are spotted contain data lost forever. According to the CD specifications, the discs should not be affected by any error of this kind.

For DVDs:

- PIE: the data in the DVD are organized in a bidimensional array with added correction codes. For each bidimensional array an error correction block is created. The Parity Inner Errors (PIE) correspond to the number of lines in which an error is detected. According to the DVD specifications there can be a maximum of 280 PI errors for every 8 consecutive ECC blocks.
- POE: the Parity Outer Errors (POE) correspond to the number of unrecoverable errors of PO codes within a ECC block. In presence of this kind of error an area of the disc is unreadable. The specifications don't allow the presence of this kind of error.

The Jitter: used both for CDs and DVDs, represents the temporal variation or imprecision in a signal compared to an ideal reference clock. It is a measure of how well defined the pits and lands of a disc are. For CD discs, jitter is defined in nanometers (nm), and the CD specification states that jitter should not exceed 35 nm. For DVD recordable discs, jitter is defined in percentage points, and should not exceed 9 %.

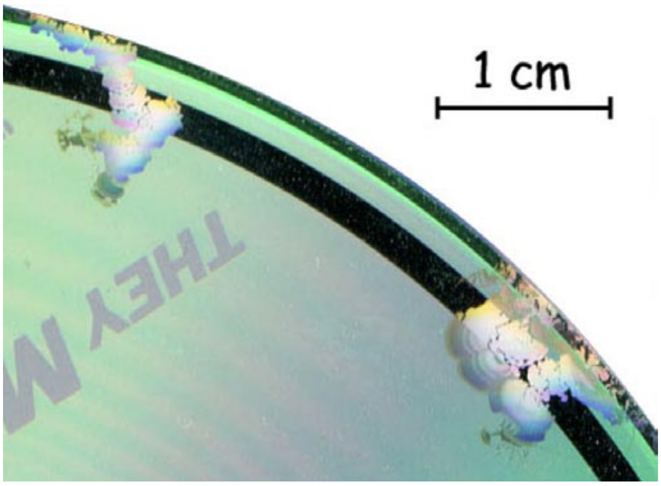


Figure 6. CDROT and rolling.

References [7][36][37][38] point out that disks with a Phthalocyanine dye and a coating composed of gold present a higher life expectancy. There exists also a silver-gold blend with the purpose of relieve the compatibility problems described in [36], with the added value of a reduction of cost. From reference [7] we take the results of the analysis presented in table III.

The tests show how the BLER of sample S4 is lower during the artificial aging tests with a Metal-Halide ligth and with extreme temperature/humidity conditions. Also the jitter and E32 errors keep at safe levels.

Most DVD's contain a stabilized Cyanine dye and it is more difficult to identify the disc stabily from the type of dye. Therefore for DVD's the selection of the more reliable discs must be done by considering samples from different manufactures,

V. Discussion

The copy of a CD or a DVD is a task that is needed for the preservation of the information contents, even if sometimes this operation is made difficult by the anti-copy systems that protect the supports. The copy of a DVD or of a non audio CD is a precise task: once the copy process is terminated, we have a disc containing the same information of the original. However this is not true in the case of an audio CD. A lot of documents [26][27][28][29] show how it is hard to get a perfect copy of an audio CD, even if it is in good condition. This is due, as we have seen, to the different strategy used to address the information. Even if these errors are often not audible, they propagate and add to others from copy to copy. A basic technique, described in [26], to check for errors generated during the copy is to take an old CD and rip its tracks, that is to read the contained musical tracks and save them into a WAV format file. The use of a WAV file is encouraged because this file format preserves all audio information without any compression. By executing this procedure using different reading devices it is possible to notice that we don't always obtain the same information.

In order to verify the error production during the ripping phase we can select an audio CD and extract the same tracks with different readers. Appropriate software is then be used to compare byte by byte each track and highlight the differences.

This phenomenon makes particularly hard to ensure a precise archive consistency over time in presence of audio CD's.

TABLE III
THE CD-R SPECIMENS FOR LIGHT EXPOSURE TEST

Sample	Coating and Dye
S1	Unknown, Super Azo
S2	Unknown, Phthalocyanine
S3	Unknown, Super Azo
S4	Silver + Gold, Phthalocyanine
S5	Silver, Metal stabilized cyanine
S6	Silver, Phthalocyanine
S7	Silver, Phthalocyanine

A constant monitoring job is therefore mandatory in order to prevent data losses: a CD or a DVD starts to decay not when it is partially unreadable but when the reader has to rely massively on redundancy codes in order to rebuild the correct information.

There are several devices on the marked, more or less expensive, that can be used to monitor the state of CD's and DVD's [30][31][32]. Moreover software solutions are available which run on the usual CD/DVD readers [33][34][35]. However the performance and the level of detail of the anlysis data differ greatly with the type of equipmnet. Therefore the choice of the right solution must be evaluated by balancing several factors such as the type of analysis required, the time dedicated to data acquisition, the data precision and the cost of the device.

In summary, the life expectancy of a CD or DVD can be very different, but however limited in time, and depends mainly on the type of disc, the type of writer, and the storage conditions. A periodical check of the level of reading errors is required to assure that back-up copies are made before the contents become unreadable.

REFERENCES

[1] Jerome L. Hartke, "Duplication Quality Tips", second edition, Media Sciences Inc, 1999.

[2] Fred R. Byers, "Care and Handling of CDs and DVDs: A Guide for Librarians and Archivists", copublished by Council on Library and Information Resources and National Institute of Standards and Technology, 2003.

[3] Rothenberg J., "Ensuring the longevity of digital documents"; Scientific American, pp. 42-47, 1995.

[4] Joe Iraci, The Relative Stabilities of Optical Disc Formats, International Journal for the Preservation of Library and Archival Material, 2005.

[5] P. Buonora, F. Liberati, F. Pizzi, F. La Camera; OptiMa Optical Media Analysis, awaiting publication.

[6] Chandru J. Shahani, Basil Manns and Michele Youket; Longevity Of CD Media Research At The Library Of Congress; Preservation Research and Testing Division Library of Congress, Washington DC, USA.

[7] Oliver Slattery, Richang Lu, Jian Zheng, Fred Byers, and Xiao Tang, Stability Comparison of Recordable Optical Discs - A Study of Error Rates in Harsh Conditions, Journal of Research of the National Institute of Standards and Technology, 2004.

[8] ECMA-119, Volume and File Structure of CDROM for Information Interchange, 2nd edition, 1987.

[9] ECMA-130, Data Interchange on Read-only 120 mm Optical Data Disks (CD-ROM), 2nd edition, 1996.

[10] Ken C. Pohlman; Principles of digital audio (fifth edition), pp. 249-314, 415-486; McGraw-Hill.

[11] ECMA TR/71, DVD Read-Only Disk - File System Specifications, 1998.

[12] ECMA-267, 120 mm DVD - Read-Only Disk, 3rd edition, 2001.

[13] ECMA-268, 80 mm DVD - Read-Only Disk, 3rd edition, 2001.

[14] ECMA-272, 120 mm DVD Rewritable Disk (DVD-RAM), 2nd edition, 1999.

[15] ECMA-273, Case for 120 mm DVD-RAM Disks, 1998.

[16] ECMA-279, 80 mm (1,23 Gbytes per side) and 120 mm (3,95 Gbytes per side) DVD-Recordable Disk (DVD-R), 1998.

[17] ECMA-330, 120 mm (4,7 Gbytes per side) and 80 mm (1,46 Gbytes per side) DVD Rewritable Disk (DVD-RAM), 3rd edition, 2005.

[18] ECMA-331, Cases for 120 mm and 80 mm DVD-RAM Disks, 2nd edition, 2004.

[19] ECMA-338, 80 mm (1,46 Gbytes per side) and 120 mm (4,70 Gbytes per side) DVD Re-recordable Disk (DVD-RW), 2002.

[20] ECMA-359, 80 mm (1,46 Gbytes per side) and 120 mm (4,70 Gbytes per side) DVD Recordable Disk (DVD-R), 2004.

[21] ECMA-382, 120 mm (8,54 Gbytes per side) and 80 mm (2,66 Gbytes per side) DVD Recordable Disk for Dual Layer (DVD-R for DL), 2008.

[22] Alberto Gaetti, tesis degree - Aspettative di vita dei supporti ottici di memorizzazione, Università Degli Studi Di Modena E Reggio Emilia - Facoltà Di Ingegneria Di Modena, 2008.

[23] ISO/IEC 10995:2008 - Test method for the estimation of the archival lifetime of optical media, 2008.

[24] Direction Des Archives De France; Recommandations relatives a la gravure, a la conservation et a l'evaluation des CD-R, 2005.

[25] A. Svensson, CD-CATS SA3 Users Manual, AudioDev Inc., West Des Moines, Idaho, USA (2000).

[26] www.exactaudiocopy.de/en/index.php/other-projects/dae-quality.

[27] www.daefeatures.co.uk (visited on 08/26/08).

[28] www.hydrogenaudio.org (visited on 08/26/08).

[29] wiki.hydrogenaudio.org (visited on 08/27/08).

[30] www.audiodev.com (visited on 08/27/08).

[31] www.eyecote.com (visited on 08/26/08).

[32] www.cd-genie.com (visited on 08/26/08).

[33] www.k-probe.com (visited on 08/28/08).

[34] www.plextools.com (visited on 08/28/08).

[35] www.cdspeed2000.com (visited on 08/28/08).

[36] Daniele Mongiello, Silvio Salza, Considerazioni sui supporti ottici, CNIPA, 2006.

[37] www.audioholics.com/education/audio-formats-technology/cd-and-dvd-longevity-how-long-will-they-last (visited on 09/04/08).

[38] www.itl.nist.gov/iad/894.05/loc/index.html (visited on 09/04/08).

Using Matroska EBML for Right Management

Cesare Bornaghi {cbornaghi@crema.unimi.it},
Ernesto Damiani{damiani@dti.unimi.it}
Dipartimento di Tecnologie dell'Informazione
Università degli Studi di Milano
26013 Crema - Italy

***Abstract-* The Matroska container is a structured system to organize multimedia data in a single file, containing audio/video streams, content tagging and video subtitles.**
Matroska is a good target for digitize audio material because multimedia data could be tagged with a limited effort (simply by generating the container).
A Matroska DRM-aware reader could be used to organize a search engine based on content search and to determine if user is entitled to access the resource.
In this paper we propose to use ad-hoc DRM computation to grant users selective access to different sections of a single Matrroska file. This approach permits to manage payment system, charging the user according to the multimedia information he actually displayed or played.

I. Introduction

Digital Preservation projects should have a target format for distribution and retrieval of digitized multimedia content.
The well-known Matroska environment [1] (http://www.matroska.org) is useful for managing storage of multimedia information, enabling advanced multimedia information retrieval.
Matroska defines an extensible open standard Audio/Video container. As of September 2008, the matroska playback downloads have exceeded 4 millions.
Matroska aims to become the standard of multimedia container formats. It was originally derived from a project called MCF, but differentiates from it significantly because it is based a binary derivative of XML called EBML (Extensible Binary Meta Language). EBML provides significant advantages in terms of future format extensibility, without breaking file support in old parsers.
Matroska audio/video container is an envelope for which there can be many audio, video and subtitles streams, allowing the user to store a complete movie or CD in a single file.
Matroska incorporates all features one would expect from a modern container format, including fast seeking in the multimedia file, advanced error recovery, selectable audio streams, which are stremable over the Internet (HTTP and RTP audio & video streams).
Matroska is an open standards project. This means it is free for personal use and the technical specifications describing the bitstream are open.
The source code of the libraries developed by the Matroska Development Team is licensed under GNU L-GPL. In addition to that, there are also free parsing and playback libraries available under the BSD license, for commercial software adaption.

The founders of Matroska have the following goals:

- Create and document a modern, flexible and cross-platform Audio/Video container format, in combination with an open codec API to form a free and open media framework
- Establish Matroska as the opensource alternative to existing containers such as AVI, ASF, MOV, RM, MP4, MPG
- Develop a comprehensive set of tools for the creation, editing and implementation of Matroska files
- Develop libraries and tools for software developers to be able to support Matroska in their applications
- Support the adoption and of Matroska's libraries into Haiku (OpenBeOS) Mediakit and GStreamer (Multimedia Framework for Linux , equivalent to Microsoft (TM) DirectShow (R) for Windows (TM))

Most of these golas have been successfully achieved. A still open problem for handling large multimedia data archives encoded in Matroska format is providing selective access to authorized section of these multimedia files.
This access control requirement springs up from two main necessities:

a) Allowing selective access for selective payment
b) Allowing selective access to enforce privacy and security policies.

The first requirement is tailored to digital archives coming from the digitization of analogical audio and video sources. It approach could be applied to digital video editing, where, for example, a single section of an AAC audio file is tagged as a DRM section and the producer is not obliged to pay rights for the whole opera but only for the section he needs to produce the media.

The second requirement is needed in law enforcement, e.g. for digital preservation of audio evidence in trials. During a trial, which might last months or even years, CDs containing wire tapping could be damaged, lost or stolen. For this reason the best thing to do is to digitize recorded material, store it in a central and secure place, make regular backup, and provide selective access.
Police officers, judges, attorneys may be granted selective privileges to different sections of digital audio material, preserving the media and avoid many inconveniences.
In this paper, we shall consider the use of the Matroska format for the distribution of multimedia archives, showing how the Matroska format can support complex DRM policies.

II. Related Work

For a lot of people the acronym DRM only means that big companies want to take control of what/when/how they watch copyrighted content [2]. Most of the time, the restrictions put on the user are considered as unfair and strongly limitating the use of legally purchased content [3].

The most common implementations of DRM are FairPlay (Apple iTunes/iPod) and Microsoft Janus (PlayForSure). The CSS protection of DVDs is also a DRM. DivX, RealNetworks and others also offer some kind of DRM.

With the possible exception of Apple FairPlay, all of the above DRM solutions can be licensed for use in a product. But for some reason, there is no way to transfer content from one format to the other. Also, make controls on moving/ copying/ selling/ lending content is impossible.

AACS, a new guideline for DRMs, will change that in the future with the introduction of Managed Copy. As the name implies, that means that the user will be able to copy DRMed content to another device as long as it does not alter the DRM rights given to the user.

While this initiative looks a very good idea, there is no guarantee that technology companies and distributors will enable this feature any time soon.

Several researchers have put forward the idea of regulating access to media based on information stored in media's content information tree (e.g. the XML trees used by many media formats). In this kind of approach information stored in XML tree is presented in a selective way, using XSL, XSLT or other type of style sheet technologies, to filter out nodes the user is not entitled to see [4] [5] [6].

Our approach, instead, relies on a container to manage and encapsulate different kind of multimedia data in an homogeneous wrapper that permits fast access and supports different cryptography systems.

The Matroska container permits to wrap audio and video files in a single container and to define a suite of tagged information by the use of the EBML.

Currently, there is no IETF endorsed MIME type for Matroska files. However it is possible to use the ones that Matroska developers have defined:

- mka : Matroska audio `audio/x-matroska`
- mkv : Matroska video `video/x-matroska`

Encryption in Matroska is designed in a very generic style that allows developers to implement whatever form of encryption is best for them. In Section VI we shall use Matroska encryption framework as a base to implement DRM.

In Matroska it is possible to manipulate encrypted streams without decrypting them. The streams can be copied, deleted, cut, appended, and a number of other editing techniques can be applied to them without ever decrypting them.

Encryption can also be layered within Matroska. This means that two completely different types of encryption can be used, requiring two separate keys to be able to decrypt a stream.

III. Units

A possible architecture using Matroska is organized around a central server where data can be stored and encrypted.

Access rights are managed via file system management access. The architecture relies on the distributed File System AFS (Andrew File System) [7].

AFS uses Kerberos [8] [9] for authentication, and implements access control lists on directories for users and groups. Each client caches files on its local file system for increased speed on subsequent requests for the same file. This also allows a limited residual file system access in the event of a server crash or a network outage.

Read and write operations on an open file are directed only to the locally cached copy. When a modified file is closed, the changed portions are copied back to the file server. Cache consistency is maintained by a mechanism called *callback*. When a file is cached the server makes a note of this and promises to inform the client if the file is updated by someone else. Callbacks are discarded and must be re-established after any client, server, or network failure, including a time-out. Re-establishing a callback involves a status check and does not require re-reading the file itself.

A consequence of this file locking strategy is that AFS does not support large shared databases or record updating within files shared between client systems. This was a deliberate design decision based on the perceived needs of the university computing environment.

A significant notion of AFS is one of *volume*, a tree of files, sub-directories and AFS mount points (links to other AFS volumes). Volumes are created by administrators and linked at a specific named path in an AFS cell. Once created, users of the filesystem may create directories and files as usual without concern for the physical location of the volume. As needed, AFS administrators can move that volume to another server and disk location without the need to notify users; indeed the operation can occur while files in that volume are being used.

AFS volumes can be replicated to read-only cloned copies. When accessing files in a read-only volume, a client system will retrieve data from a particular read-only copy. If at some point that copy becomes unavailable, clients will look for any of the remaining copies. Users of that data are unaware of the location of the read-only copy; administrators can create and relocate such copies as needed. The AFS command suite guarantees that all read-only volumes contain exact copies of the original read-write volume at the time the read-only copy was created.

The file name space on an AFS workstation is partitioned into a *shared* and *local* name space. The shared name space (usually mounted as /afs on the Unix filesystem) is identical on all workstations. The local name space is unique to each workstation. It only contains temporary files needed for workstation initialization and symbolic links to files in the shared name space.

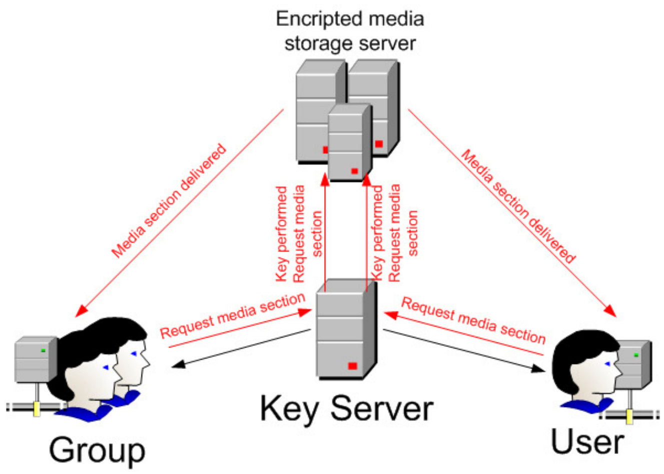


Figure 1. Architecture of the system

In the case of Matroska files, system access is managed by the usual user and groups authentication by credentials.
Access to each single section of the media is provided by the use of a decryption jkey served by the key server during the authentication phase.
The key provided is an encrypted hash computed using the user and group public keys. This key contains an hash that has a one to one correspondence to the EBML tag into the encrypted file system. This correspondence permits to download only the *temporal section* and *content section* of the media matching the hash.
The correspondence is computed by server decoder that allows clear rendering just for sections enabled by the hash content of the key.
This is permitted by the separation of layers ensured by the Matroska specification.

Figure 2. Matroska information specification.

The EBML layer is separated from content layer; this decoupling makes it possible to manage multimedia information in fast way.
In **Figure 3** a definition of matroska data container is shown.
Let us summarize what each segment contains.

1. The Header contains information saying what EBML version this files was created with, and what type of EBML file is this. In our case it is a Matroska file.

2. The Metaseek section contains an index of where all of the other groups in the file are located, such as the Track information, Chapters, Tags, Cues, Attachments, and so on. This element isn't technically required, but you would have to search the entire file to find all of the other Level 1 elements if you did not have it. This is because any of the items can occur in any order. For instance you could have the chapters section in the middle of the Clusters. This is part of the flexibility of EBML and Matroksa.

3. The SeekID contains the "Class-ID" of a level 1 element. The Meta Seek section is usually just used when the file is opened so that it can get information about the file. Any seeking that happens when playing back the file uses the Cues.

4. The Segment Information portion gives us information that is vital to identifying the file. This includes the Title of the file and a SegmentUID that is used to identify the file. The ID is a randomly generated number. It also has the ID of any file that should be associated with it.

5. The Track portion tells us the technical side of what is in each track. The name of the track goes in Name. The tracks number goes into the TrackNumber element. And the TrackType tells us what the track contains, such as audio, video, subtitles, etc. There are also settings to tell us what language is it in, and what codec to use for playback of the track. Each Track has a unique ID called TrackUID, much like the ID for the whole file. This can be used when you are editing files and have several different versions, it makes it easy to see what files have that specific track. The TrackUID is also used in the Tagging system.

6. The Segment Information section contains basic information relating to the whole file. This includes the title for the file, a unique ID so that the file can be identified around the world, and if it is part of a series of files, the ID of the next file.

7. The Track section has basic information about each of the tracks. For instance, is it a video, audio or subtitle track? What resolution is the video? What sample rate is the audio? The Track section also says what codec to use to view the track, and has the codec's private data for the track.

8. The Chapters section lists all of the Chapters. Chapters are a way to set predefined points to jump to in video or audio.
9. The Clusters section has all of the Clusters. These contain all of the video frames and audio for each track.

10. The Cueing Data section contains all of the cues. Cues are the index for each of the tracks. It is a lot like the MetaSeek, but this is used for seeking to a specific time when playing back the file. Without this it is possible to seek, but it is much more difficult because the player has to 'hunt and peck' through the file looking for the correct time code.

11. The Attachment section is for attaching any type of file you want to a Matroska file. You could attach anything, pictures, webpages, programs, even the codec needed to play back the file.

12. The Tagging section contains all of the Tags that relate to the file and each of the tracks. These tags are just like the ID3 tags found in MP3. It has information such as the singer or writer of a song, actors that were in the video, or who made the video.

Figure 3. Matroska deep content description.

IV. INFORMATION MANAGEMENT USING AUTHENTICATION

The Extensible Binary Markup Language EBML was designed to be a simplified binary version of XML for the purpose of storing and manipulating data in a hierarchical form with variable field lengths. Specifically EBML was designed as the framework language for the video container format Matroska. Some of the advantages of EBML are:

- Compatibility between different versions of binary languages defined in EBML. A rare property of binary format that otherwise often needs careful consideration beforehand.

- Unlimited size of data payload.

- It can be both generated and read as a stream, without knowing the data size beforehand.

- Often very space efficient, even compared to other binary representations of the same data.

EBML has also some well-known disadvantages:

- No references can be made between EBML files, such as includes or inherits. Every EBML document is a self contained entity. The data stored in EBML may of course reference other resources.

- No compositioning process to merge two or more EBML files currently exists.

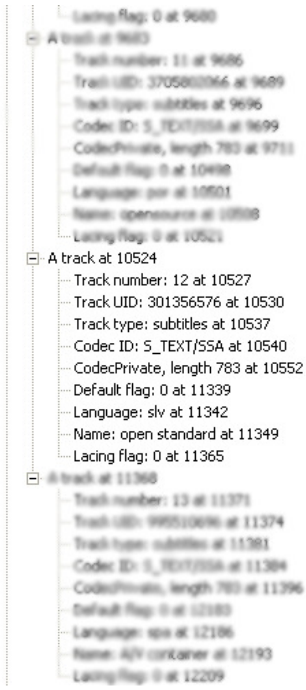


Figure 4. EBML example of readable content

For our purpose, a fundamental features of EBML is that it supports differential encryption.
For example on a single video .mkv, using EBML is possible to read just one track because other track are obfuscated by the cryptographic algorithm.

The track "10524" is accessible using selective reading zone defined by the hash included in the authentication key.
Other section of the file are not accessible if the key does not permit the access.

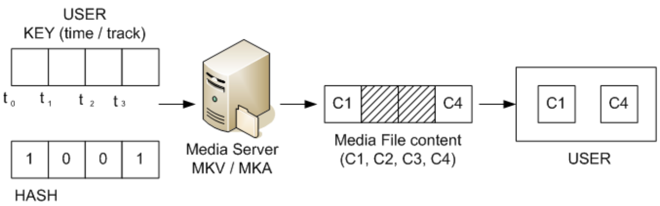


Figure 5. Complete structure of media interchange.

Is important to underline that information transmitted on the net are just the information resolved by the user key performing. In this kind of approach there is a minimization of data that could be stolen.

V. SOME EXAMPLES

Audio content tagging could be organized using Matroska, simply generating the container.
For example an audio file called *"audiofile.aac"* could be encapsulated in an .mka audio container using mkvmerge GUI [1], which is a graphic tool used to manage information and multimedia data in a Matroska file.
Just one file is merged in the container in **Figure 6**. The output file is defined by the path shown in the box called "Output filename". Usually, only for audio media files it is suggested to use .mka container, but .mkv is also used.

Adding single tag is possible using "Global" tab menu and inserting the tag into "File/segment title".
After that, it is necessary to choose if the tag has to be associate to the duration of a temporal segment or to a size in Mbyte of a section of the file or after a specific timecode.
It is also possible to associate a single title to different audio files already merged into the same container.
The last thing to do is to start muxing the file/files and information to merge in the new .mka file with the button "Start muxing" shown in **Figure 7**.

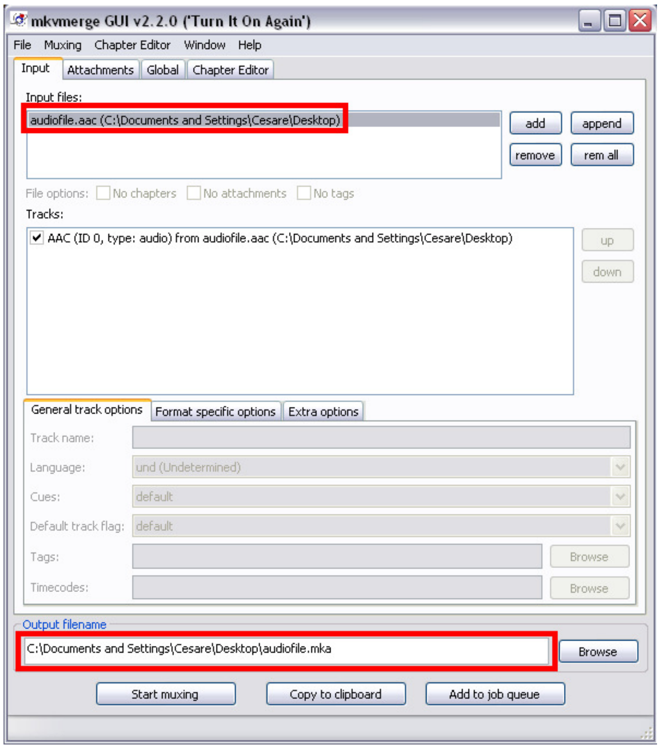


Figure 6. Opening an audio file using mkvmerge GUI.

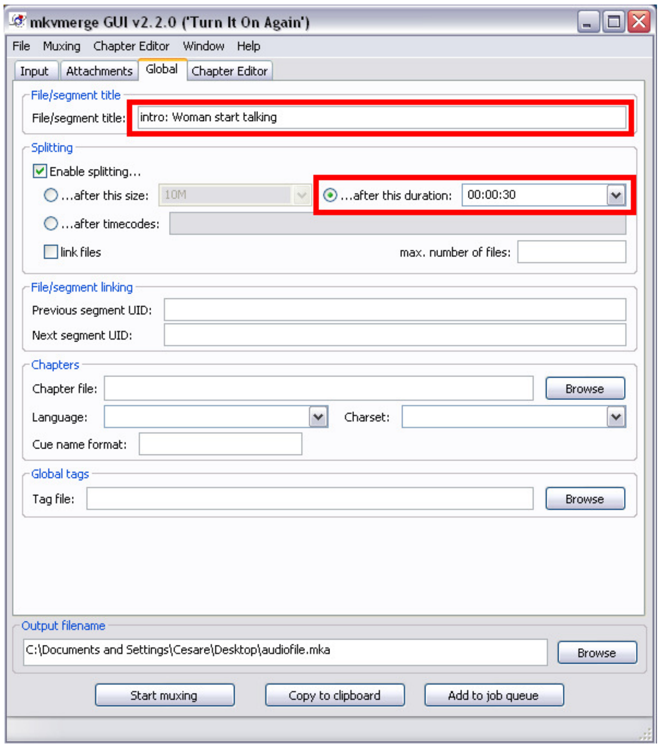


Figure 7. Tagging audio file adding segment title to a single track

Let us now examine how a simple DRM policy can be enforced based on EBML tagging. Tags are presented by the user accessing the file and are parsed by resolver to search matching to EBML content tags into Matroska file. For instance these keys can be obtained when performing AFS authentication.

The policy allows the user to access a section of ExampleAACMedia.aac. Resource identifier is 00148 and track identifier is 3832603748. The Media container holds only audio format in A_AAC/MPEG2/LC codec standard. User is authorized access the recording via hash key given by key server. The hash identifier is *asfdusafda545a6sas54sdc*.

The user can only read the media because write and modify access are not permitted by the policy.

After the parser and a resolver running on the server have done the match between the XML policy (containing hash key), and the Matroska File EBML tags, the section of the media is encoded in a new file with a temporary name and sent to the client.

```
<Policy>
  <Resource>
 <ResourceName>ExampleAACMedia.mka</ResourceName>
    <ResourceDescrition>
      AAC Media record for example policy
application.
    </ResourceDescrition>
    <ResourceID>00148</ResourceID>
    <ResourceSection>
      <TrackNumber> 2 </TrackNumber>
         <TrackUID> 3832603748 </TrackUID>
         <TrackType> audio </TrackType>
          <CodecID> A_AAC/MPEG2/LC </CodecID>
         <Duration> 21.333ms </Duration>
    </ResourceSection>
  </Resource>
  <User>
    <Name> Cesare <Name>
    <Access> LocalDownload </Access>
    <UserID> 3832603749 <UserID>
    <HashID> asfdusafda545a6sas54sdc </HashID>
  </User>
  <Privileges>
    <read> yes </read>
    <write> no </write>
    <modify> no </modify>
  </Privileges>
</Policy>
```

The resulting EBML tags coming after policy enforcement are shown below through the use of the `mkvtoolnix` tool:

```
|  + Track number: 2
|  + Track UID: 3832603748
|  + Track type: audio
|  + MinCache: 0
|  + Timecode scale: 1.000000
|  + Codec ID: A_AAC/MPEG2/LC
|   + Default duration: 21.333ms (46.875 fps for a
video track)
|  + Language: und
|  + Audio track
|   + Sampling frequency: 48000.000000
|   + Channels: 2
|  + Default flag: 1
```

VI. Conclusion

In this paper we have shown how the modular structure of a Matroska container can support the enforcement of DRM policies.

Matroska can therefore be seen as therefore a key target for the DRM-aware distribution of multimedia content coming from digitization projects.

References

[1] Matroska. Extensible open standard Audio/Video container. URL:www.matroska.org

[2] Method and system for sychronizing audio and visual presentation in a multimodal content renderer URL:www.freepatentsonline.com

[3] Joseph Thomas-Kerr, Ian Burnett, Christian Ritz *Enhancing Interoperability via Generic Multimedia Syntax Translation*, 2006, Proceedings of the Second International Conference on Automated Production of Cross Media Content for Multi-Channel Distribution (AXMEDIS'06)

[4] Naren Kodali, Duminda Wijesekera *Regulating access to SMIL formatted pay-per-view movies*, 2002, Proceedings of the 2002 ACM workshop on XML security

[5] E. Damiani and S. De and Capitani Vimercati and E. Fernández-medina and P. Samarati, *An access control system for SVG documents*, King's College, University of Cambridge, UK, 2002, pages: 29—31, Kluwer Academic Publisher

[6] Ernesto Damiani, Cristiano Fugazza, *Toward semantics-aware management of intellectual property rights*, Online Information Review, Emerald, 2007, pp. 59-72

[7] John H. Howard. *An overview of the Andrew file system*. In Proceedings of the USENIX Winter Conference, pages 23--26, Berkeley, CA, USA, January 1988. USENIX Association.

[8] J.G. Steiner, C. Neuman, J.I. Schiller, *Kerberos: an authentication service for open network systems*, in: USENIX Conference Proceedings, Dallas, TX, Winter 1988.

[9] Steiner, J. G., Neumann, B. G., and Schiller, J. I. *KERBEROS: An authentication system for open network systems*. In Proc. Winter

Modeling and structuring data for cultural heritage: the Pinakes 3.0 project and the deployment of standard

Andrea Scotti
Fondazione Rinascimento Digitale
Firenze

Abstract—**This article aims to present mainly three basic things: a) the results of a three year activity of the Pinakes Project. Therewithin will be exposed how such tool can serve the scholarly community and the general public working in the humanities and in the cultural heritage[1]; b) how it is possible to crossover different typologies of data set gathered from different disciplines; c) how its possible with this application to work on distributed data both at a generalist level and at a specialist one. Finally we will show how data, such as those coming from the sound management have been treated within this project.**

I. STRUCTURE OF THE APPLICATION[2]

The entire application is build only with codes and side application based on the Open Source[3] policy. Such application is granting the possibility to furnish at the end user a dynamic interface to model, insert/publish and navigate data. In order to achieve such not easy task has been developed a logical model which generalization could offer the method to represent any given knowledge domain based on shared foundational categories. The implication depending of the abstraction can be explained as follows: do not offer a descriptive model but a method with which is possible to generate descriptive models of reality. In such a way the user, that has clear what is needed in his knowledge domain, applying the Pinakes 3.0 methodology, can decide how to represent and relate the objects of his reference context.

In order to reach such a level of generalization has been developed what in the computational world is called foundational ontology. The latter is a tree representation of the undeletable classes needed to represent any entity of reality. Entity is here intended as follows: any object of experience both physical or logical of which is needed a semantic description.

The classes

Therefore we have separated all reality into four different basic classes: the physical object (res extensa), the logical object (any entity of reality non having an extension nonetheless object of experience), semantic object and digital object. These four classes, starting out from the root called element are a specialization of the super class OWR (Object World Reification). Beside such main classed we have also define a set of minimal attributes controlled by a number of service classes dependent from a super class called RWR (Related World Reification). All these classes can be specialized but not deleted. Finally there is a class devoted to the control of the primitive values (integer, Boolean, string, lob).

Let's see in detail the main classes. The physical object has a minimal number of attributes in order to describe any real entity of which are given the following information: the extension, the matter/s, and the property. In fact no object of which is given an extension has a mater and is property of someone. This rule is true from the museum object to a landscape. The logical object at the foundational level contains only the descriptive name of the object it self. This class is a container of semantic objects i.e. denotators that can refer to concepts, processes, non extensive phenomena. Both the physical and the logical objects exist only on the basis of a given semantic object that denotates its content. Independently of the typology of description (generic or specialized) its denotation is the only condition of its cognoscibility. Therefore the semantic object contains the time, space, responsibility, and anthroponomy.

The super class quoted above RWR manages these attributes. We may specify better the content of such classes:

- *Anthroponomy* that has as sub class physical person, hero, can be specialized if it needed to group for example all

[1] This project started in 2006 and is supported by FDR (Fondazione Rinascimento Digitale) in Florence, the Institute for Computational Linguistic of the CNR (National Research Council) in Pisa and is promoted both from the Istituto e Museo di Storia della Scienza in Firenze and the Italian Misnistry for Cultural Heritage

[2] For a more detailed history and the theoretical background of the application Pinakes 3.0 see: Andrea Scotti, Pinakes: Structuring and Destructuring Documentation in the Humanities. A Project for Modelling Data in History Research. In: Michael Stolz, Lucas Marco Gisi u. Jan Loop (Hg.), Literatur und Literaturwissenschaft auf dem Weg zu den neuen Medien. Bern, germanistik.ch 2005 (Literaturwissenschaft und neue Medien). This article is downloadable at: http://www.germanistik.ch/publikation.php

[3] See http://opensource.org. Form now on OSI (Open Source Initiative)

mythological gods and find out where they are geographically born;

- *Toponym*, that is using the space classification suggested by the Getty Museum [4];

- *Time* and *time representation*, to describe time ranges and also their calendar representation;

- *Minimal attributes of the semantic object*, that manages the serial relation existing between: person → its responsibility → time → time reference → place →;

- *Digital objects* the standard required attributes to describe any object machine-readable or produces by machines

II. EXAMPLES

In order to have an idea how all this works it is needed to show some snap spots from the application currently distributed. The first image shows the foundational ontology as we have described it above. Such basic schema is the only not editable part of the application and grants the interoperability among projects or projects groups. In fact any possible specialization done on the ontological schema is transparent on the base of the common super classes. These are so to speak model classes to create real classes. Any new sub class will inherit the attributes of the super classes. It goes without saying that to introduce new classes of a given knowledge domain its is required to have both the knowledge of the discipline and that of the Pinakes methodology.

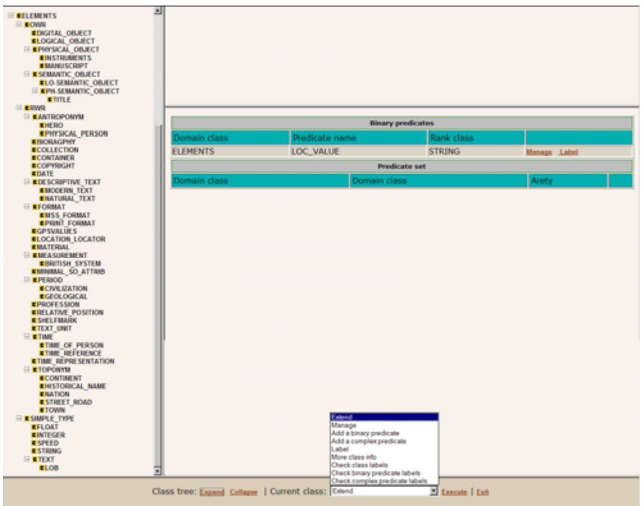


Fig. 1 The expanded three class of the foundational ontology

[4]See: http://www.getty.edu/research/conducting_research/vocabularies/tgn/ particularly the geo-database costumed on the art history and archeology.

This is the schema that the user will have working in the administration area of the application. In this example the physical object class has been specialized with the classes instruments and manuscripts. The specialization for the semantic object class is the class title. These classes can be defined by the users and have, above the attributes that inherit from the abstract classes, new ones depending on the need of the project.

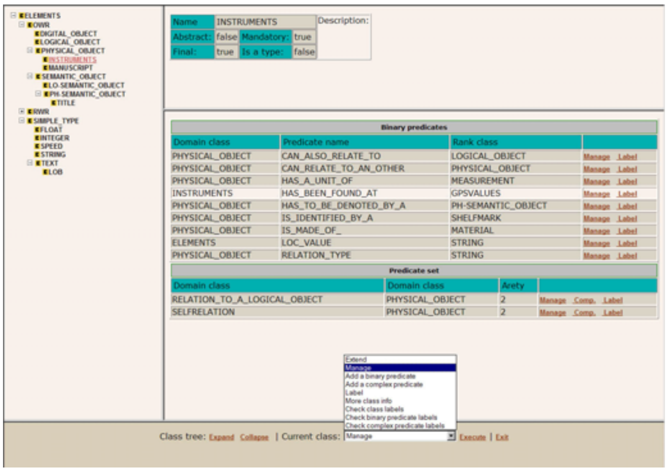


Fig. 2 Specialization of a class of the basic schema

In order to make an example a new attribute has been created (the line with white background). This attribute, named has_been_found sets into relation a given instrument with its geo-referenced values. To carry out this operation we have to manage the basic class (physical object), recognizable from the grey background, adding and defining a binary predicate connecting the domain instrument with the range GPS.

The GSP class has been added to the service classes' area RWR (see picture.1). These new attributes will be read from the input application that will return a new input field for each new attribute added.

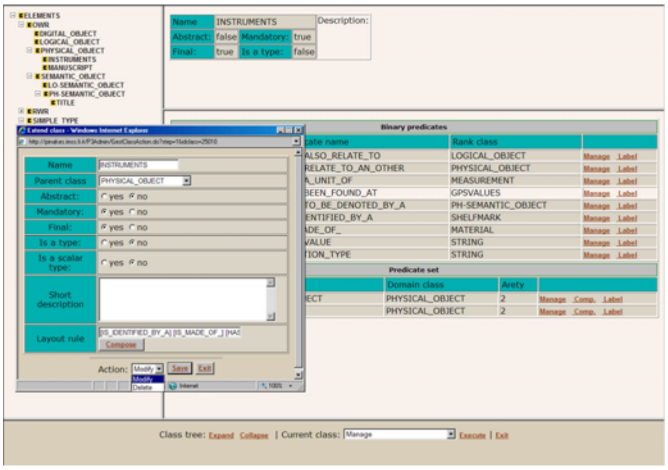


Fig. 3 Class administration: specialization

To carry out successfully the operation described above we have to understand what types of statements are required to define a new sub class. In short, following the label on the left of the image: the class name; if the class is compulsory for our schema or not; if the class can be specialized or not; if the class is type meaning a service class; if the values of the class a scalar meaning if the values are alphanumeric strings or not; description of the class content; and finally in which way we want to have the short form of the class instances.

Let's see now how the input form will visualize our classes. At first a drop down menu will show us all the super classes so that we will be obliged to choose one starting from there to add/edit our data.

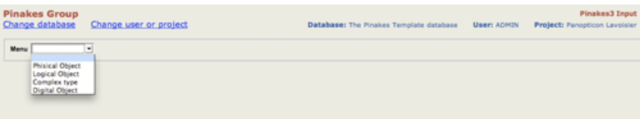


Fig. 4 (a) Choice of a generic object

We will choose the physical object and we will have the chance to insert the date following the order and typology we have decided in the schema.

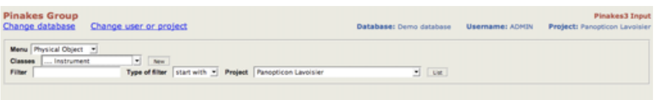


Fig. 4 (b) Choice of particular object

Here we will find the class list we have defined on the schema. The data input /edit can start or clicking on the button new or searching data to be edited using the box filter.

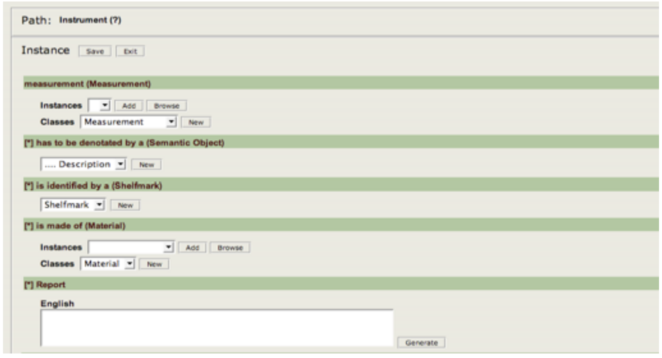


Fig. 4 (c) The form to insert new data

The input interface uses the rules defined on the schema to determine which kind of functions should be assigned to each field (pull down menu, access to controlled vocabulary etc.) This dynamic method to generate the interface give the user a chance to model it's own date also during the research activity. The labelling in fact could be drawn from the disciplinary taxonomy and in any case is defined by the user. The only part of the entire application that should not undergo any kind of transformation is the foundational ontology. In fact by changing that structure you bring out of synchrony your data. In order to see, use and being seen and used the common element is the foundational ontology.

III. CONCLUSIONS

Finally, we believe that the effort undertake till today has successfully shown that is possible to create applications that the user can use to define its own domain of knowledge being able then also to store, edit, publish it working exclusively on the web. The idea that within the humanities you can have a consortium of data distributed over the different communities, is still something not experimented. Nonetheless, we hope that ideas and methods on which we have build such an application will be a basis for the development of that discipline known under the name of Computing & Humanities

REFERENCES

[1] Andrea Scotti, Pinakes: Structuring and Destructuring Documentation in the Humanities. A Project for Modelling Data in History Research. In: Michael Stolz, Lucas Marco Gisi u. Jan Loop (Hg.), Literatur und Literaturwissenschaft auf dem Weg zu den neuen Medien. Bern, germanistik.ch 2005 (Literaturwissenschaft und neue Medien). This article is downloadable at: http://www.germanistik.ch/publikation.php

[2] OSI (Open Source Initiative)http://opensource.org.

[3] D. Oberle, Semantic Management of Middleware. Semantic Web and Beyond, Springer, 2006.

[4] W. McCarty, We would know how we know what we know: Responding to the computational transformation of the humanities, 1999. httl://www.cch.kcl.ac.uk/legacy/staff/wlm/essays/know

[5] K. de Smedt e.a. (Eds.), Computing in Humanities Education. A European Perspective, Bergen, University of Bergen, 1999 (Socrates/Erasmus Thematic Network Project on Advanced Computing in the Humanities). http://gandalf.aksis.uib.no/AcoHum.

Assessing long term preservation of audiovisual digital contents with DRAMBORA

Perla Innocenti [1], Andrew McHugh [2] , Seamus Ross [3] , Raivo Ruusalepp [4]
[1,2,3] Humanities Advanced Technology and Information Institute (HATII) at the University of Glasgow
[4] Nationaal Archief , Netherlands

Abstract

The Digital Curation Centre (DCC) in the UK and the EU-funded DigitalPreservationEurope (DPE) project jointly released the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA, http://www.repositoryaudit.eu/) in early 2007, with the goal to provide a practical, evidence-based toolkit for assessing repositories and digital libraries. Subsequent iterative development has let to the refinement of its methodology, and the release of DRAMBORA Interactive, a freely available online tool aimed at streamlining the core risk assessment process. DRAMBORA represents a bottom-up approach that takes risk and risk management as its principle means for determining digital repositories' success and for charting their improvement. The tool's development and ongoing evolution has been informed at all times by practical research. More than twenty international repositories have been subject to assessment using DRAMBORA, enabling the validation of its primary methodology and offering insights into potential shortcomings and the extent of its applicability in a range of diverse preservation contexts. Furthermore, these exercises have enabled initial research into repository profiling, which attempts to identify commonalities within subsets of the repository community in order to inform and facilitate subsequent repository development and evaluation. This paper describes the DRAMBORA methodology, focusing on its benefits and developments, and introduces DRAMBORA Interactive. It goes on to describe the results of some of the most successful pilot assessments. Most notable is the work funded by the DELOS Digital Library project, which sought to identify core characteristics within a range of textual and audiovisual digital libraries, in order to conceive a repository profile that might form the basis for subsequent repository development and evaluation

Keyworkds: Digital preservation; digital curation; risk assessment; audits; digital repositories; digital libraries; audiovisual content

1. Audiovisual / multimedia content, digitization and digital libraries

Within non-book materials, audiovisual documents are probably the most difficult to be defined. According to IFLA [1], the term "audiovisual" is related to audio and/or vision and audiovisual materials include every type of audio and/or still or moving images. The concept of multimedia adds further complexity to this definition, but it might also be more appropriate. Multimedia indicates a representation of reality which adopts diverse communication media, and might include one or more audiovisual expressions. From a technical point of view, audiovisual materials can be produced with diverse techniques, on a wide variety of media and are subjected to various manipulations.

For textual and still images, digitizing an item and preserving the original chromatic, graphic and dimensional attributes is a conceptually simple challenge. Currently available technology is not longer confined to research labs, but is commercially available at a professional level. Nevertheless, digitization technology brings along a number of issues that do not have immediate solution: lack of skills of content holders to perform the digitization themselves, possible cost increment of a third party specialized digitization service, heterogeneity of materials which requires the definition of multiple digitization methodologies and tools, difficulty in defining the use, integration and structured access modalities of digital content. A proper digitization project is frequently missing: technicians are often exclusively interested in developing algorithms, and users often confuse automation with technology.

Furthermore, as technologies allow such conversion, it is fundamental to identify reference standards for the original materials, rather than standards referring to the digital system or the planned output. It is also essential to define the selection of use for digitization technologies and devices, to choose the best quality within a predefined range conducting a benchmarking of the characteristics of available products.

Audiovisual and multimedia materials encounter the same difficulties, but with a further level of complexity. In theory, digitization allows a lossless reproduction of the original audiovisual quality. It simplifies post-processing operations (material restoration and re-use), automatic or semi-automatic extraction of metadata, and use of Multimedia Information Retrieval systems [2]. But in reality, the restoration of a audiovisual content such a film is an ambiguous operation, a reproduction on a diverse media. Cinematographic restoration, for instance, has got nothing to do with painting restoration. The tight bond to the original support belongs to the arts, where nothing else except our visual apparatus is needed to make them accessible, while analogue audiovisual content always require a technological device. Moving from an analogue to a digital film, the visual perception is remarkably different. But digital format allows to asynchronously working in subsequent phases (whereas with a analogue film you need to simultaneously intervene on both media). It permits the virtual reconstruction of an initial document that, as with the case of 16mm film, can be so fragile to be unlikely to be moved onto another analogue support.

In addition to the above mentioned constraints, further challenges are created by the cost of specialized devices, which require skilled staff. Those devices do not only include specific digitization devices and software, but also playback devices for both analogue and digital content. Compressing methods for audiovisual content are evolving, and uncompressed audiovisual files may require significant storage space and network bandwidth to make them accessible online. The timeframe for digitization generally is on a 1:1 scale, content editing of digitized materials is equally time-consuming and the computational power needed to process the digitized content is still towering. Legal, preserving and cataloguing criteria and standards are still evolving and can bring challenging differences at national and international level.

For all the above mentioned challenges, digital library and repository design and management for audiovisual content (which often takes into account textual and non-textual materials) can be a taxing task. As outlined in the DELOS Digital Library Manifesto [3], the digital library universe is a complex framework in which at least three types of conceptually different "systems" can be identified, namely, digital libraries (DLs), digital library systems (DLSs) and digital library management systems (DLMSs). Architecture, personalization, quality, policy and usability are essential to the design and deployment of digital libraries (and of the digital repositories at their heart). But if we cannot ensure the long-term sustainability of the content, ensuring the presence of these capabilities would be pointless. Therefore, we require mechanisms that will enable us to measure the success of digital libraries and their underlying repositories in content preservation, as this is a fundamental building block of a digital library system and environment.

2. The landscape of digital repositories assessment criteria

The contemporary domain landscape suggests that information repositories are likely to play a role of considerable importance in the pursuit of digital preservation assurances.

In order to legitimise decentralisation to smaller scale repository environments, it is essential that the community has appropriate mechanisms available to support repository assessment, and determine the competencies of those charged with information stewardship responsibilities. Management, staff, financiers and partners must all be satisfied that their efforts are capable of meeting formal expectations. Similarly, information creators, depositors and consumers naturally hope to obtain similar assurances of the capabilities of the organisations providing maintenance, preservation and dissemination services.

Considerable work has been undertaken to develop preservation audit check-lists, intended to represent the objective benchmarks against which repositories' efforts are judged. The two primary examples, both released in 2007, are:

The *Trustworthy Repositories Audit and Certification (TRAC) Criteria and Checklist* [4] describes approximately ninety characteristics that repositories that aspire to a certifiable, trustworthy status must demonstrate they have;

The *nestor Catalogue of Criteria for Trusted Digital Repositories* [5] reflects the regional needs of the nestor community. Structured similarly to the TRAC document, this provides examples and perspectives that are more representative of a German operational, legal and economic context.

Both TRAC and nestor are compelling reference materials, and their usefulness in informing the development and retrospective evaluation of repositories is widely acknowledged. However, neither is sufficient in isolation. By their very nature, check-lists like these adopt a top-down assessment philosophy: both examples seeking to define an *objective* consensus of the priorities and responsibilities that should exist within any repository environment. By relying solely on *nestor* or *TRAC*, one implicitly disregards the great variety that is visible across contemporary digital repository platforms. The question persists, is a one-size-fits-all approach to assessment and certification really useful for those within the curation community? Both *TRAC* and *nestor*'s criteria have been painstakingly phrased to ensure their flexibility, and facilitate optimal general applicability. But despite such efforts, it appears evident that within the community there is the need for a more tailored assessment solution that takes into account atypical repository qualities, as either a companion piece, or alternative, to the other existing guidelines.

The *Digital Repository Audit Method Based on Risk Assessment* (DRAMBORA) [6] developed by the Digital Curation Centre and DigitalPreservationEurope is designed to address such shortcomings. Its bottom-up approach enables repositories to relate their benchmarks for success more explicitly to their own aims and contextual environment, enabling an increased granularity of understanding of preservation approaches and challenges. Furthermore, by focusing explicitly on the process of assessment, rather than simply listing desirable repository characteristics, it provides considerably more opportunities for evidence-supported, demonstrable excellence, and consequent repository confidence. A key strength is that DRAMBORA is capable of being used both independently and in association with more objective guidelines.

3. DRAMBORA Opportunities and Outcomes

Digital curation can be characterized as a process of transforming controllable and uncontrollable uncertainties into a framework of manageable risks. The DRAMBORA process focuses on risks, and their classification and evaluation according to individual repositories' activities, assets and contextual constraints. The methodological outcome is a determination of the repository's ability to contain and avoid the risks that threaten its ability to receive, curate and provide access to authentic and contextually, syntactically and semantically understandable digital information.

DRAMBORA acknowledges the heterogeneity that exists within the digital world, refraining from explicitly describing the characteristics that repositories should demonstrate. Instead, parameters for success are aligned with the subjective mandate, objectives and activities of individual repositories. Specific contextual factors and constraints are considered only where they are relevant. This ensures that the results of the audit process are, from the participating repository's perspective, wholly applicable and immediately useful. The process aims to provide repositories with formal understanding of their own mandate and objectives, to provide them with a detailed and manageable breakdown of fundamental challenges, promote communication within the organisation as a whole and facilitate subsequent external audit whether based on TRAC, nestor or any other repository assessment criteria.

3.1 Origins and alignment with international initiatives

In 2006 and early 2007 the Digital Curation Centre (DCC) undertook a series of pilot audits in a diverse range of preservation environments. Various repositories participated, exhibiting a range of different characteristics [7]. As well as providing the participating organisations with an objective and expert insight into the effectiveness of their operation, and determining the robustness and global applicability of those metrics and criteria already conceived [8], the audits were aimed at exploring the optimal means for conducting assessment of repositories. The research set out to develop an increased understanding of how evidence can be practically accumulated, assessed, used and discarded throughout the audit process. A methodology for performing repository audit was quickly established and subjected to considerable subsequent refinement. In March 2007 the process was formalised as the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA), and a first textual version of the toolkit was released.

Important consensus about the breadth of repository characteristics that must be exposed to scrutiny during an assessment process was reached during a meeting of the authors of DRAMBORA, TRAC and nestor in early 2007. Adopting a broad view that echoed the work done by RLG/OCLC in their seminal 2002 "Trusted Digital Repositories – Attributes and Responsibilities", ten general principles of repositories were conceived. The ten principles [9] are varied, encompassing more than simply technological considerations, extending to organisational fitness, legal and regulatory legitimacy, appropriate policy infrastructures, mandate and commitment, and every aspect of object management, including ingest, preservation, documentation and dissemination. For DRAMBORA's purposes, these can be conveniently grouped according to three core criteria classifications, each influenced by contextual factors and exposed to risk, as illustrated in Figure 1.

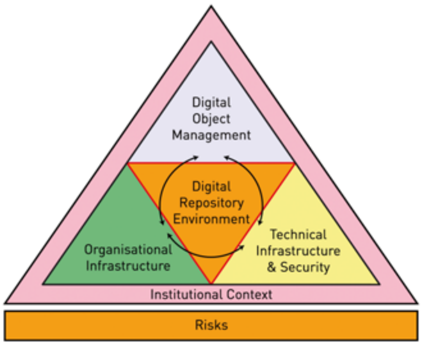


Figure 1: interrelationships within a digital repository environment. © HATII at the University of Glasgow

3.2 Methodology

DRAMBORA's approach is flexible, and responsive to the structural and contextual variety evident within textual and audiovisual repositories: its metric for success is directly linked with repositories' own aims.

Evidence and demonstrable success are at the very forefront of the DRAMBORA process. The first phase of assessment reflects this, a process of information accumulation, aggregation and documentation. The repository's strategic purpose, its action plan, and any contextual factors that influence or limit its ability to meet its objectives must each be made explicit. A hierarchical analysis is undertaken; definition of the repository's mandate is the first step of an increasingly focused scrutiny, requiring detailed descriptions of fundamental repository objectives as well as the activities intended to ensure their successful achievement. The outcome of this phase is a comprehensive organisational overview, which immediately leads into the latter phase, concerned with the identification of risk.

The issue of risk has been considered from a number of perspectives within the context of digital curation and preservation. For instance, a variety of work has sought to analyze the risks associated with particular file formats, perceiving the risk as something intrinsic to what a digital repository does, based upon the technical challenges associated with maintaining the usability of digital files and storage media [10]. More recently some authors, such as Ross [11] and Ross and McHugh [12], have described the inherent uncertainty associated with digital preservation.

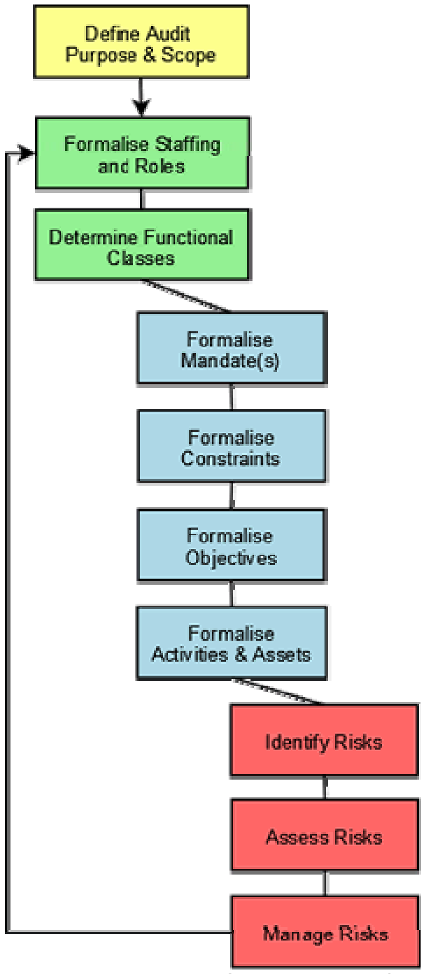


Figure 2: DRAMBORA audit workflow

The risk identification, assessment and management part of the DRAMBORA process is where conclusions are derived from the organisational picture conceived within the first phase. Risk is utilised as a convenient means for comprehending repository success – those repositories most capable of demonstrating the adequacy of their risk management are those that can have, and engender, greater confidence in the adequacy of their efforts. Preservation is after all, at its very heart, a risk management process. The fundamental temporal challenges of preservation are naturally complicated by future uncertainties. Threats relating to any number of social, semantic and technological factors are capable of inhibiting long term access to digital materials

4. DRAMBORA Interactive

In early April 2008, in response to usability issues associated with an entirely paper-based approach, a second version of the toolkit was released as DRAMBORA Interactive, a freely available web based tool (Fig. 3) [13]. DRAMBORA interactive leads auditors through the individual stages of the assessment process, recording and displaying responses and providing greater structure to facilitate a more comprehensive coverage. The tool provides robust security provisions, supporting multiple repository contributors, but protecting potentially sensitive information from non-authorised access.

The tool's implicit workflow exactly reflects the core DRAMBORA methodology. In addition, characteristics of each registered repository can be described in detailed terms, with technological, organisational and resource related issues made explicit. This facilitates the intelligent comparison of objectives, challenges and risks with those of peer repositories, again, intended to maximize the assessments' breadth of coverage. The tool is equipped with numerous reporting mechanisms to visualize the repository's status, and support the improvement planning process.

5. Digital Library Repository Profiling: the DELOS audits

DRAMBORA Interactive was primarily developed to inject greater practical usability into the assessment process, but since its development, further advantages have revealed themselves. Perhaps most notably, the developers have DRAMBORA have identified opportunities for repackaging assessment responses to provoke or inspire individuals within comparable repository contexts. Ultimately, such information will form the basis for a series of repository profiles capable of encapsulating core roles, responsibilities, functions and risks for a variety of repository types. The availability of these profiles is expected to facilitate and further legitimise both repository assessment and development. Currently, repository profiling measures correspond (but need to be limited) to the descriptive fields already utilised within the *DigitalPreservationEurope* project's repository registry [14]. By defining their own characteristics, the DRAMBORA software is thereafter equipped to offer targeted suggestions.

Some theoretical work has already indicated the feasibility of these efforts. Within the context of the DELOS Digital Preservation Cluster four audits of digital library environments were undertaken, using DRAMBORA. The Michigan-Google Digitization Project and MBooks at the University of Michigan Library, Gallica at the Bibliothèque nationale de France, the Digital Library of the National Library of Sweden and CERN's Document Server exhibit a range of organisational and functional characteristics representative of most of that which is conceivable within the digital library context. Two out of four organizations included in their holdings audiovisual material: Gallica at the Bibliothèque nationale de France (sound recordings) and CERN's Document Server (videos). Each assessment incorporated an onsite visit that took an average of three days, preceded by a lengthy period of dialogue and information exchange between project facilitators and institutional participants, and considerable desk-based research. The conclusions that followed each would be distilled into a broadly applicable generic template, focusing not on diversity, but the fundamental commonalities that distinguish and characterise digital libraries.

Applying risk analysis based auditing methodology to digital libraries has identified both common strengths and weaknesses in their work. While digital libraries are highly efficient in automating ingest of digitised content, and providing flexible access to their collections, the acquisition of born-digital content poses more difficult requirements that need bespoke solutions and often semi-automatic processing. For metadata and access digital libraries can rely on existing library standards and electronic catalogues that can be linked to simple storage solutions. Relying primarily on standard formats has to some complacency surrounding the digital preservation challenge. This is exaggerated further because digitised collections represent little more than access-facilitating surrogates of their analogue collections, and this is understandable. Each participating institution demonstrated adequate technical infrastructures, and sufficient security to maintain the digital library services.

The areas where the audited digital libraries collectively fail or show weakness relate to:

- lack of policies and procedural manuals and maintenance of the organisation's knowledge-base;
- creation and management of preservation metadata;

- documentation of the systems in use and provision of an audit trail of processing applied to digital objects in library's care;
- maintaining transparency to its stakeholders and involving them in improving the digital library services;
- delegation of responsibility for preservation planning and effective preservation strategy building.

All participating libraries were in the process of expanding and changing their services, which was expected to bring these weaknesses increasingly to the fore. The audited digital libraries can be described as risk-minimal digital repositories, and are certainly aware of their shortcomings. Hence, they were each well placed to earn the status of a trustworthy digital repository. A detailed report [15] on these audits has been published with detailed description of the audit findings.

5.1 Findings from the DELS audits

The process of assessments yielded almost as many insights about the assessment tool itself as the current state of digital libraries. A further conclusion highlights the suitability of DRAMBORA within an ever-evolving digital context. The four organisations that participated in this process are all, like the peers they represent, in a state of transition. New services are being developed, expansions are being planned to other areas, new contracts are being signed and new responsibilities embraced as novel legislation emerges. DRAMBORA metric is much more focussed on facilitating improvement than on the imposition of transitory judgements. In that respect its iterative workflow has a great deal in common with maturity modelling, which is expected to be integrated in an increasingly formal way within DRAMBORA in subsequent iterations

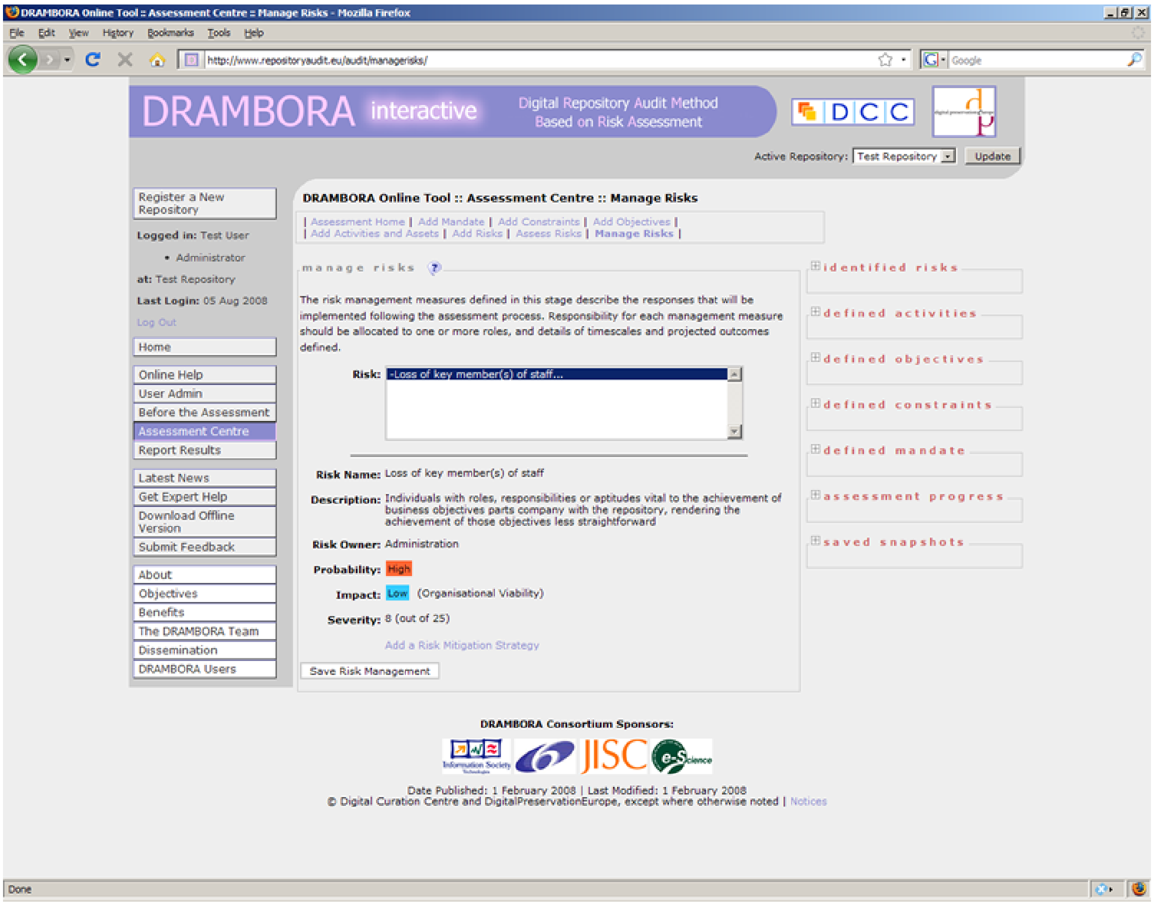


Figure 3: DRAMBORA Interactive interface: Risk management section

5.2.1 Supported self-audits

The most overwhelming response from the audited institutions was that the DRAMBORA audit process yielded numerous benefits, and provided insights that would undoubtedly prompt further investigation and probable response. However, a general response that appeared to be consistent from each of the audited organisations was the value of the process would be lessened if the DELOS facilitators were not present. This is of some concern, given DRAMBORA's role as self assessment methodology. Because of the bottom up approach favoured within DRAMBORA, within its defined self audit process the parameters for success are associated with the specific aims and mandate of the audited repository. As described above, by requiring users to describe the characteristics of their own repositories, DRAMBORA Interactive presents 'comparable organisations' with insights into the kinds of risks that are faced by their peers, in order to help ensure a more comprehensive coverage. The development of meaningful repository profiles, that reflect contextual realities of the preservation process, is expected to represent the ultimate outcome of this.

5.2.2 Staff participation to the audits. In order to be of real value to the organisation, everyone with any relevant responsibilities or concerns ought to be involved. Communication on an organisation-wide basis is always acknowledged as vital, but all too often overlooked or underemphasised. The self-audit represents an invaluable opportunity to develop a shared and globally acceptable interpretation and understanding of overall strengths, weakness, opportunities and threats. However, although a wide range of representation is vital to ensure the audit's success, this must be well managed. Representations should be planned to ensure that discussions are logistically feasible and that no more than four individuals are involved at any time. As more participants are added beyond that number the discussion will become increasingly difficult to manage, and focus more and more difficult to maintain. Conversational tangents become more common and fundamental audit questions might remain unanswered, or answered only in an incomplete or superficial sense.

5.2.3 Risk scoring. With respect to risk assessment, two priorities for repository staff emerged during the audits. The first was to build a relative array of risks, capable of illustrating where the mildest and most severe challenges within the organisation were evident; the other is to establish how the repository or digital library's maturity compares with that of its peers. The two are far from incompatible, but in order to present useful, globally comparable results the apportioned scores must have some objective significance. Descriptions of the significance of the available scores are presented within DRAMBORA, but these are not immune to further interpretation. For this reason, when DRAMBORA is utilised to support a self assessment process, its results are of most value for internal use. Involving an external (and consistent) facilitator enables these results to have considerably greater objective weight, and may then be the basis for a more global comparison.

A vital commodity when describing risk is a means to determine, or express risk impact. It appears that the perception of challenge associated with preservation within digital library contexts is quite distinct from that of those dealing with born digital or otherwise unique digital assets. In most cases within the audited institutions, the value of digital content was mainly surrogacy for physical assets. Libraries remain primarily access-focussed and digitised content is considerably more plentiful than born-digital materials. Preservation is naturally prioritised lower since, notwithstanding the significant cost of rescanning large quantities of content, anything that is lost can generally be digitised again. An objective risk impact scoring system that considers only one manifestation of success or failure is unnecessarily restrictive. Consequently, the risk impact expressions have now been overhauled within DRAMBORA, so that for any risk auditors can select the terms within which impact is realised. A weighted model has been favoured, with four 'risk expression' types which can be scored according to a common scale. Irrespective of the specific practical units with which risk impact might be quantified (e.g., in Euros, Gigabytes or a less tangible measure), the impact is described uniformly. The new impact expressions are:

- Reputation and Intangibles
- Organisational Viability
- Service Delivery
- Technology.

These are assumed to be proportionate loss areas, but individual responses can reflect priorities that are adopted by auditing institutions. Impact continues to be measured according to a scale from very low to very high, although the interpretative text that accompanies each has been neutralised to support any of the four risk impact classifications, and permit comparability.

6. Conclusion

DRAMBORA has now been deployed in a range of evaluative contexts, and the processes of self assessment and facilitated assessment continue to yield considerable insights into both preservation activities, and the state of preservation assessment. Work associated with DRAMBORA will continue a variety of ways, from training activities to international audits and collaborations. The developers of DRAMBORA have or have had active collaborations with the following international initiatives and projects: Trustworthy Repository Audit and Certification (TRAC) Criteria and Checklist Working Group, Center for Research Libraries (CRL) Certification of Digital Archives Project, Network of Expertise in Long-term storage of Digital Resources (nestor), DELOS Digital Preservation Cluster (WP6), International Audit and Certification Birds of a Feather Group,

SHAMAN (Sustaining Heritage Access through Multivalent ArchiviNg).

The DCC and DPE are committed to training a generation of DRAMBORA auditors through a number of planned events taking place in 2008 and 2009. Facilitated audits will continue both interactively and through physical visits, with new organisations registering their repositories and completing self-assessments every week. DRAMBORA Interactive was released in early 2008 and the procedure to submit DRAMBORA as the basis of an ISO standard has been initiated (ISO TC46 /SC 11). DPE and the Digital Curation Centre intend to continue to develop DRAMBORA to support the longer term management of repositories and ensuring that they are auditable and continue to develop in ways that enable them to consistently improve their levels of service and the longer term sustainability. They will also support its widest possible take-up within the United Kingdom, Europe and broader international contexts.

Acknowledgements

The development of DRAMBORA has been enabled by the European Union through its Framework Six support for DigitalPreservationEurope under the IST Programme (IST-2005-2.5-10) through grant IST-2005-034762. The DigitalCurationCentre (DCC) in the UK was the other key funder of DRAMBORA development

References

[1] "Audiovisual and Multimedia Section. Guidelines for Audiovisual and Multimedia Materials in Libraries and other Institutions", [Online document], March 2004, IFLA, Available HTTP http://www.ifla.org/VII/s35/pubs/avm-guidelines04.htm.

[2] Perla Innocenti, Roberto Raieli, Il contesto del MultiMedia Information Retrieval (MMIR), in L'informazione multimediale dal presente al futuro: le prospettive del MultiMedia Information Retrieval, AIB-Lazio, 2005, pp. 101-140.

[3] A digital library is "a (potentially virtual) organization that comprehensively collects, manages, and preserves for the long term rich digital content and offers to its user communities specialized functionality on that content, of measurable quality, and according to prescribed policies" (DELOS Digital Library Manifesto, p. 8, [Online document], DELOS, 2006, Available HTTP: http://www.delos.info/index.php?option=com_content&task=view&id=345

[4] TRAC was developed by a consortium jointly overseen by the US National Archives and Records Administration and the Research Libraries Group (prior to its absorption within OCLC), and is now maintained by the Center for Research Libraries. Available HTTP: http://www.crl.edu/PDF/trac.pdf.

[5] The nestor Catalogue developed in Germany by the Network of expertise in Digital long-term preservation. Available HTTP: http://edoc.hu-berlin.de/series/nestor-materialien/8/PDF/8.pdf.

[6] DRAMBORA toolkit, DRAMBORA Interactive and more information on ongoing activities are available at http://www.repositoryaudit.eu/ . See also A. McHugh, R. Ruusalepp., S. Ross and H. Hofman, Digital Repository Audit Method Based on Risk Assessment, [Online document], Digital Curation Centre (DCC) and DigitalPreservationEurope (DPE), 2007, Available HTTP: http://www.repositoryaudit.eu/.

[7] The British Atmospheric Data Centre (BADC); the National Digital Archive of Datasets (NDAD); the National Library of New Zealand's National Digital Heritage Archive (NDHA); the Florida Digital Archive (FDA) at the Florida Centre for Library Automation; and the Beazley Archive (BA) at the University of Oxford were among those that kindly agreed to take part. For a mechanisms to roll out audit and certification services for digital repositories in the United Kingdom, see S. Ross, A. McHugh, "The Role of Evidence in Establishing Trust in Repositories", D-Lib Magazine, July/August 2006, Volume 12 Number 7/8, Available HTTP: http://www.dlib.org/dlib/july06/ross/07ross.html.

[8] Most notably, at that stage, the RLG/NARA Draft Audit Check-list for Certification of Digital Repositories, which would latterly be published as TRAC.

[9] "Core Requirements for Digital Archives", [Online document], Center for Research Libraries (CRL), 2007, Available HTTP: http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=92.

[10] Risk Communication Tool, ERPANET (2003), Available HTTP: http://www.erpanet.org/guidance/docs/ERPANETRiskTool.pdf. Cornell University Library Virtual Remote Control (VRC) tool, Cornell University, Available HTTP: http://irisresearch.library.cornell.edu/VRC/methods.html. JISC, Managing Risk: a Model Business Preservation Strategy for Corporate Digital Assets, (2005), Available HTTP: http://www.jisc.ac.uk/whatwedo/programmes/programme_preservation/programme_404/project managingrisk.aspx. G. Lawrence, W. Kehoe, O. Rieger, W. Walters and A. Kenney, Risk Management of Digital Information: A File Format Investigation, CLIR Report no. 93 (2000), Available HTTP: http://www.clir.org/pubs/reports/pub93/pub93.pdf. V. Lemieux, Managing Risks for Records and Information, ARMA International (2004). N. McGovern, A. Kenney, R. Entlich, W. Kehoe and E. Buckley, "Virtual Remote Control. Building a Preservation Risk Management Toolbox for Web Resources", D-Lib Magazine, vol. 10, no. 4, April 2004, Available HTTP: http://www.dlib.org/dlib/april04/mcgovern/04mcgovern.html

[11] S. Ross, Uncertainty, Risk, Trust and Digital Persistency, NHPRC Electronic Records Research Felloships' Symposium Lecture, University of North Carolina at Chapel Hill (2006)

[12] S. Ross, H. McHugh, "The Role of Evidence in Establishing Trust in Repositories", D-Lib Magazine, July/August (2006), vol. 12, no. 7/8 (Also published in *Archivi e Computer*, August 2006), Available HTTP: http://www.dlib.org/dlib/july06/ross/07ross.html.

[13] Available HTTP: http://www.repositoryaudit.eu/.

[14] Available HTTP: http://www.digitalpreservationeurope.eu/registries/repositories/.

[15] S. Ross, A. McHugh, P. Innocenti, R. Ruusalepp, Investigation of the potential application of the DRAMBORA Toolkit in the context of digital libraries to support the assessment of the repositories aspects of digital libraries, DELOS Noe for HATII at the University of Glasgow, 2008.

Workshop on Cultural Heritage and Artificial Intelligence

Organized by

ENEA Centro Ricerche (Italy)

Coordinated by

Luciana Bordoni, ENEA C.R. (Italy)

AXMEDIS 2008: proceedings of the 4th International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution. Workshops, Panels and Industrial Applications. Florence, Italy 17-19 November 2008
edited by Paolo Nesi, Kia Ng, Jaime Delgado
ISBN: 978-88-8453-811-6 (online). ISBN 978-88-8453-810-9 (print).

Sonification of narrative texts: an experimental study

Roberto Basili, Federico Spini, Saverio Evangelista
University of Roma "*Tor Vergata*"
Via della Ricerca Scientifica, snc, 00133 Roma (Italy)
Email: basili@info.uniroma2.it, fede@gmail.com, s_evangelista@email.it

***Abstract*—This paper presents an on-going experimental activity in the area of musical composition. In particular, a framework for the support of automatic composition based on a stochastic model applied to families of FM synthesizers has been fully developed. The novelty is that the entire process is parametrized according to semantic similarity and relevance measures as derived from the tracking of semantic/narrative concepts in a literary opera.**

I. Introduction

> "...the qualification *beautiful* or *ugly* makes no sense for sound, nor the music that derives from it; the quantity of intelligence carried by the sounds must be the true criterion of the validity of a particular music."
>
> - Iannis Xenakis ([1]

Linguistic creativity is a unique property of human language. It manifests in our ability to combine known words in a new sentence, in the expression of thoughts in figurative languages, and in general, as a support for creative usages of different communication processes even on the visual or musical dimensions. A specific use of linguistic creativity is the automatic musical composition inspired by semantic analysis of texts. This task consists in the translation of complex concepts, as they manifest in a literary text, into a musical piece. We are interested to methodological and technological aspects of this process and a study carried out over the novel "Gli Indifferenti" by Alberto Moravia is here presented.

The objective of the study is the development of a musical artifact as the outcome of the analysis of a text, something that corresponds to a process of *sonic* interpretation. In the literature, such process is often called *sonification*. The input to this process is a semantic description of lexical meanings as they manifest in a text, in particular in a literary opera. As we will see (in II-A) this description is a quantitative semantic model with a geometric nature. It correspond to a notion of *closeness* in a space between points that represent concepts: the distance between them is interpreted as a form of *semantic relevance*. The linearity of the original literary text gives rise to a syntagmatic interpretation of this model as it evolves linearly along the text. Values of relevance change across the text and they give naturally rise to numerical sequences that represent the evolution of the corresponding concepts in the text.

The problem involved by our notion of sonification is the adoption of these numerical sequences as the effects of an underlying semantic process. If sequences are the visible outcomes of a stochastic process the sonification can be realized as the computation of their most likely *explanation*. In line with other language processing tasks ([2]), the observable phenomena here are the word sequences (or better sequences of their semantic relevance in texts) and they correspond to emissions of a generative device that is in a given state. In our work, state changes in the devices are made corresponding to changes in the literary and semantic content of the opera. A decoding process can be here used to compute the most likely state sequence able to optimally justifying the observed phenomena. This approach has analogous either with early approaches to automatic composition (see [3]) and with current NLP methods, where Hidden Markov Models (HMM [4]) are largely adopted to efficiently solve the decoding problem. A relationship is imposed with sonification whenever we assign a musical meaning to each state. Individual states are mapped into specific musical actions so that state sequences give rise to manipulation of computable representation of musical objects, such as instruments, synthesizers or MIDI sequences. As the actions are *evoked* by the observable semantic changes in the text, the musical composition is here the side-effect of an interpretation of semantic phenomena. This is what we call here *sonification*.

In Section II a linear model of lexical semantics and the ways it can be derived from a text are discussed. We will then discuss the architecture and the techniques of the proposed sonification process (Section III). A case study is then briefly reported in Section IV.

II. Extracting Semantics from Texts

Lexical meaning is at the basis of most tasks in Information Retrieval, such as ad hoc retrieval, document clustering or summarization. Although the formalization of word meaning is an old topic in AI and Philosophy, IR approaches have traditionally approached this huge problem, by relying on simple meaning surrogates, i.e. the words themselves, with an extraordinary success, in terms of accuracy and scalability, given the shallow nature of the adopted representation. When using lexicalized features (such as the words occurring in a text to express the latter's semantics), several advantages arise. First, all the observations of the proper features are objectives and errors in the data interpretation are avoided[1]. Discrete, although large scale, feature sets can be naturally mapped into possibly high-dimensional vector space representations, where geometrical metrics supply principled real-valued functions as models of semantic similarity. Finally, analytical methods for manipulating the derived space can be inherited from the huge tradition of linear algebra and optimization theory. Notice how this is especially relevant in the text processing area where lexical features belong to large scale dictionaries (e.g. even millions of features are observed within Web collections), and the dimensionality curse is critical for realistic (e.g. Web) applications. The well known distributional hypothesis suggests that word meaning can be acquired through a wittgensteinian "*language in use*" perspective and the growing availability of collections of digital documents allows to explore it on a large scale. It has been recently observed that distributional models allow to acquire in rather inexpensive ways several forms of lexical information: from topical associations (e.g. *doctor* vs. *nurse*) to paradigmatic (*in absentia*) information (e.g. *doctor* vs. *professor*) to syntagmatic knowledge [5].

[1]Notice how this is not true when external knowledge, e.g. semantic networks such as Wordnet, or syntactic analysis is applied for the acquisition of more complex features

A. Latent Semantic Analysis

Studies on learning methods for pattern recognition and automatic classification tasks have outlined the role of geometric transformations for dimensionality reduction ([6], [7], [8]). These aim at capturing the subset of significant information implicit in the data distribution itself, and representing this source information by means of the minimal number of dimensions. Although several applications have already demonstrated the impact of these methods on the reachable accuracy and the scalability guaranteed by reduced models, the full implications on lexical acquisition and modeling tasks have not been fully explored yet.

Latent Semantic Analysis (LSA) is an algorithm presented by Deerwester et al. in [9], and afterwards diffused by Landauer [6]: it can be seen as a variant on the Principal Component Analysis (PCA) idea. LSA aims to find the best subspace approximation to the original document space, in the sense of minimizing the global reconstruction error projecting data along the directions of maximal variance.

LSA captures term (semantic) dependencies by applying a matrix decomposition process called Singular Value Decomposition (SVD). The original term-by-document matrix M, that describes traditional term-based document space, is transformed in the product of three new matrices: U, S, and V such that $M = USV^T$. Matrix M is approximated by $M_k = U_k S_k V_k^T$ in which only first k columns of U and V are used, and only first k greatest singular values are considered. This approximation supplies a way to project term vectors into the k-dimensional space using $Y_{terms} = U_k S_k^{1/2}$ and document vectors using $Y_{docs} = S_k^{1/2} V_k^T$. Notice that the SVD process accounts for the eigenvectors of the entire original distribution (matrix M). tightly dependent on a global property. The original statistical information about M is captured by the new k-dimensional space which preserves the global structure. Each dimension (i.e. an induced LSA features) may be thought of as an artificial concept and represents emerging meaning components from many different words and documents [6].

B. LSA and narrative texts

LSA represents a paradigm alternative to logical and meta-linguistic approaches to meaning (e.g. predicative structures in generative linguistics or logical formalisms for ontology representation and reasoning). LSA pushes for an analytical and geometrical view on meaning within a Vector Space Model paradigm. The concepts emerge from texts, as a consequence of a similarity metrics grounded on the relations among texts and lexical items. Figure 1 depicts an example of LSA-based space where regions express word clusters as emerging concepts: in the example, a context[2] of a word, *bank*, is shown as a point in the LSA space. Its surrounding includes other lexicals like *river*, *hill* or *gate* that naturally trigger the proper "river bank" sense of *bank* and characterize the micro-domain of the source sentence.

Distance in the latent semantic space gives rise to a natural notion of *semantic domain*[10], [11], fully expressed on a lexical basis. *Semantic Domains are clusters of terms and texts that exhibit a high level of lexical coherence. They are also characterized by sets of domain words, which often occur in texts about the corresponding domain*" ([10]).

[2]"*He took a walk along the bank of the river*"

The idea here pursued is to exploit the automatic acquisition of semantic domains as a form of intelligent support to the critical interpretation of narrative texts. It is worth to be noticed that *knowledge related to target literary work* itself can be represented through domains determined only by the opera and the relations there emerging. Moreover, a structured LSA-based analysis is possible.

Narrative analysis is usually fed with the collocational evidence as it is found in the target texts. However, structured knowledge about a novel is not directly realized in atomic lexicalized phenomena, i.e. word occurrences. For example, when studying a work like "*Gli Indifferenti*" by Moravia [12], the notion of *noia* (*boredom*) is central to the analysis of some of the novel's characters. It is not straightforward to capture such structured notion only by means of simple atomic lexical information, i.e. words. Notice that the word "*noia*" itself is not so frequently used in the novel (it appears just 18 times and is the 634-th words in the frequency ranking). Second most of its morphological variants (e.g. *annoiare* (*to bore*), *annoiato* (*bored*)) are not captured collocationally with *noia*. Third collocational analysis has no way to capture most of its topically associated words, like *esistenza* (*existence*), *avventura* (*adventure*), falsita' (*falsity*, *pretence*), ...

Semantic domains can be captured from extended corpora making use of the semantic distance established in the latent semantic space, a further type of analysis of each domain can be directly done against the opera itself. Notice how each textual unit of the opera is a sort of pseudo document and is also represented as a point into the LSA space. Again distance in this space can be assumed here as a narrative information. Semantic similarity (the dual notion of semantic closeness) suggests how much a textual unit t (e.g. a paragraph) is related to a semantic domain c, i.e. at what extent a critical analysis of the target opera should take t in consideration as an embodiment of the notion c.

Moreover, textual units are either individual paragraphs or entire chapters of the book. They are strictly ordered and give naturally rise to a syntagmatic view on the target narrative work. In this way

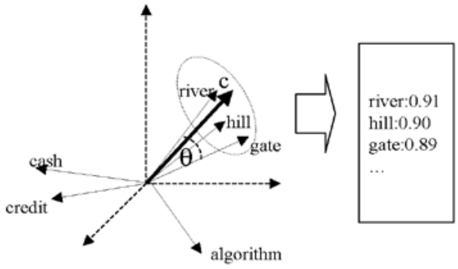


Fig. 1. An example of geometric representation of semantically associated lexicals

similarity can be established not only locally but as a dynamic notion that proceeds across individual units and follows the narrative development. Notice that whenever a quantitative notion of similarity among a narrative concept and a paragraph (a text portion enclose between the beginning and the ending of the line) is available the narrative development can be expressed graphically as its function along the totally ordered set of units. A graphical expression of a complex semantic domain is thus achievable and can be made fully available operationally.

The above two aspects require from one side an expressive definition of a semantic distance (or dually of a similarity function). On the other side an additional model that sees the opera as a sequence of possibly structured units is needed. So, paragraphs will be assumed as atomic notions. Chapters are sequences of paragraphs so that similarity at the level of chapters is an aggregation function of the similarity function over individual paragraphs. Finally the entire opera can be seen as a sequence of chapters. The graphical metaphor can depict similarity along the linearly organized sequence of chapters, or along the sequence of paragraphs internal to a chapter. An analysis at different degrees of granularity is thus made possible.

C. A quantitative model for narrative concepts

The semantic distance function, adopted for this stage of the analysis, is defined as the cosine similarity within the LSA space generated over the opera. In this context, given a concept c, its lexicon as derived from a larger external corpus C_E, and given a textual unit t of the original opera, i.e. $t \in C$, the semantic similarity between c and t is the cosine similarity among their vectors, as they are represented in the LSA space generated over the only opera, i.e. LSA_C. More precisely,

$$sim(c,t) = cos\_sim(\vec{c},\vec{t}) = \frac{\sum_i c_i t_i}{||\vec{c}||||\vec{t}||} \quad (1)$$

where $\vec{c} = \sum_{w \in L(c)} \vec{w}$, $\vec{t} = \sum_{w \in t} \vec{w}$ and c_i t_i are the i-th components of the vectors ($\vec{c}$, $\vec{t}$). $\vec{\cdot}$ is here always to be intended as the representation in the LSA_C space[3]. Notice how the lexical items in $L(c)$ are derived from an LSA-based analysis in the extend corpus C_E. Here their representation restricted to C is used, so that LSA_C is intended.

We can see here the opera T as a sequence of textual units t_i. A quantitative representation of the narrative development of c in a text can be then obtained by a discrete function $f : \aleph \times T \rightarrow \mathcal{R}$, where $\aleph$ is the abstract space of narrative concepts. f can be defined as follows:

$$f(t_i, c) = \frac{sim(c, t_i) - \mu}{\sigma} \quad (2)$$

where t_i represents the i-th unit of the opera, μ and σ are the mean and standard deviation values of the $sim(c, t_i)$ distribution, respectively. Here different distribution can be assumed with respect to the locality adopted. Different grains can be targeted so that the mean (or standard deviation) can be obtained over a chapter Ch_i (by averaging across

[3] $\vec{w}_i$ is obtained by multiplying the i-th row of the original term-document matrix with the mapping matrix $TS^{1/2}$, derived according to the SVD transformation.

paragraphs $t_j \subset Ch_i$) or over the entire opera T (i.e. by averaging across chapters $Ch_j \subset T$) The plot of the function $f()$ provides a graphical representation of the behavior of the relevance of c across different groupings of textual units in the entire opera, i.e. paragraphs or chapters.

Given a chapter $Ch \subset T$, as a subsequence of length $n < N$ of the original $T = t_1, ..., t_N$, the overall semantic similarity between Ch and c requires an aggregation function Ψ the maps individual contributions local to paragraphs into a global score. More precisely, given a narrative concept $c \in \aleph$ and a chapter Ch, the similarity function among the two is given by:

$$f(c, Ch) = \Psi_{t_i \in Ch}(f(t_i, c)) = \frac{1}{N} \sum_{t_i \in Ch} f(t_i, c) \tag{3}$$

Equation 3 expresses the aggregation as the standard mean value of the discrete distribution of values $f(t_i, c)$. Experimental evidence as acquired from the analysis of "*Gli Indifferenti*" by Alberto Moravia ([12]) will be discussed in Section IV.

In order to determine a lexical description for a semantic domain related to a notion of narrative interest (like *noia*) is derived according to three textual resources:

- relevant words defining the domain, that we call *cue words*, e.g. *noia*
- critical texts associated to the opera that we call extended corpus, C_E
- the opera itself C characterized by paragraphs and chapters as *textual units* t

The LSA space model is build on the matrix words-paragraphs that insist on the extended corpus C_E or on the opera C, that will be referred LSA_E and LSA_C, respectively. It models the relationships between words and their contexts within different knowledge levels. It allows to capture a particular semantic domain of a cue word triggered by a specific text portion.

Now, in order to characterize a semantic notion, like *noia*, we have two possibilities: study its behavior in the opera (i.e. over the corpus C) or associating to its discussion the bundle of social evidence also given by critical reviews of the opera itself (i.e. over the corpus C_E). Notice how the first choice is tight to the author's view on the concept, that is the fact and the narrative evidence intrinsic to the work, this including people, events and locations, discussed in the text. This view can be partial as it does not capture the implicit role of readers that make reference to a wider evidence, i.e. their experience and knowledge of the world. The adoption of the extended corpus augment the generalization power of the system as it may refer to every situation (i.e. piece of textual evidence) available. This is more general and expressive of the overall semantics underlying the target narrative concept suggested by the cue word.

The adopted process can thus be formally expressed as follows. Given a cue word c and the extended corpus C_E:

- First, run LSA on to the C_E and make available the transformation matrices $TS^{1/2}$ and $S^{1/2}D^T$ that map term and document vector in the transformed LSA_E space.
- Map the cue word c in the LSA_E space, i.e. compute the vector $\vec{c}$ in the LSA space
- Select words w that are close enough in LSA_E to the cue word c, as the lexicon L_c characterizing the semantic domain $S(c)$. More precisely

  $$L_c = \{w \in C_E \text{ such that } ||\vec{c} - \vec{w}|| < \tau\}^4 \tag{4}$$

  where $||.||$ is the cosine similarity distance in LSA_E and τ is a positive constant aiming to control the generalization trade-off, required not to introduce too much noise in the process.

Equation 4 defines mathematically the notion of neighborhood of a word, as depicted in Fig. 1, aiming to capture lexical cohesion among topically related words of a semantic domain. The result of the above process is a subset of the overall dictionary L_c (i.e. words belonging to the opera as well as words used in the critical reviews) that characterize the semantic domain of the cue word c.

[4] It should be noted here that threshold τ can be made dependent on individual words w, so that words more relevant for the corpus are given some preference. Technically in our experiments, $\tau_w = \frac{\tau}{ln(tf_w)}$ where tf_w is the term frequency of w.

For example from the cue word *noia*, we obtained $L_{noia} = \{$*esistenza*, *atto*, *avventura*, *fatalita'*, *falsita'*, *familiare*, *ragazza*, *abitudini*, ...$\}$[5].

A comparison of the domains (*noia* and *automobile*) is shown in Fig. 2, where the values of the corresponding $f(c, Ch_i)$ derived from Eq. 3 are reported for all the 16 chapters of the novel "Gli Indifferenti". Oppositions are outlined in the figure, as the relevance of the two domains expresses different semantic implications across the chapter sequence.

III. SONIFICATION AS STOCHASTIC COMPOSITION

A system for *sonification* is targeted to the following major tasks:

1) Generation of musical events as actions over some main features (such as duration, pitch and timbre)
2) Ordering of these events according to some principle or schemata
3) Event compilation as the final audio rendering of sequences of actions

As every musical event acts on a specific computable representation, models of actions for the different formats are required. Audio, symbolic representation (such as scores or MIDI) or synthesis ([13]) undergo different actions and must be properly abstracted. We call them *Sonic Types* (ST). A sonic type is an abstraction of a simple musical object (e.g. a short audio sample) that can be sequenced, combined and manipulated during a composition process. Events acting over these abstractions are clearly dependent on the nature and format of the involved (ST) objects. They also constraint the of possible actions on them that we call *Action Types*. Action types are classes of functions able to modify individual Sonic Types. The automation of a sonification process thus requires the solution of the following tasks:

- The definition of a set of Sonic Types on which a composition can act;
- The design of computationally-tractable abstractions for Sonic Types;
- The definition of parametrized functions, the Action Types, suitable to manipulate Sonic Type instances providing abstractions for the musical actions required during a composition

While a sonification platform is here required to govern an wide variety of Sonic Types, each individual composition is based on a specific set of their instances that will characterize the sonification results. Choices about format (e.g. MIDI sequences vs. oscillators), timbral or pitch aspects are here defined *a priori*, as a reference protocol by the composer, according to different artistic objectives. In our system the source material for a composition is specified through the so-called *Sonic Dictionary* (SD). A SD describes the SonicType and the specific parameters of each musical device involved in a specific composition. Typical parameters that characterize for example the modulating oscillators in an FM generator (i.e. the Sonic Type FMgenerator) are here basic frequencies and waveforms: these are specified as features for the FMgenerator Sonic Type and their instances are declared into the Sonic Dictionary of a musical piece. Correspondingly, Sonic Types trigger some potential Action Types. Every Sonic Type can be manipulated through some actions that allow to express individual musical events on the corresponding instances.

An overall view of the sonification process is reported in Figure 3. Equations 3 and 4 express a complex model of a literary concept c as it manifests in a text. Given the total order of the sequences of $f(c, Ch)$ values we can interpret them as observables from an emitting device, i.e. the opera. In Fig. 3 we refer to this process as the extraction stage. What we still need is a computational model of the generation phase, the core step in our sonification process. A pure object oriented approach has been here used where Sonic Types, Action Types have ben mapped into Java objects, and a declarative language for defining the Sonic Dictionary has been developed. These are able to abstract away from the specific implementation details that characterize different musical materials, such as MIDI vs. FM synthesizers, and allow to transparently support the generation process, as a stochastic interpretation process.

[5] $L_{noia} = \{$*existence*, *act*, *adventure*, *deceit*, *falsity*, *relative*, *girl*, *habits*, ...$\}$

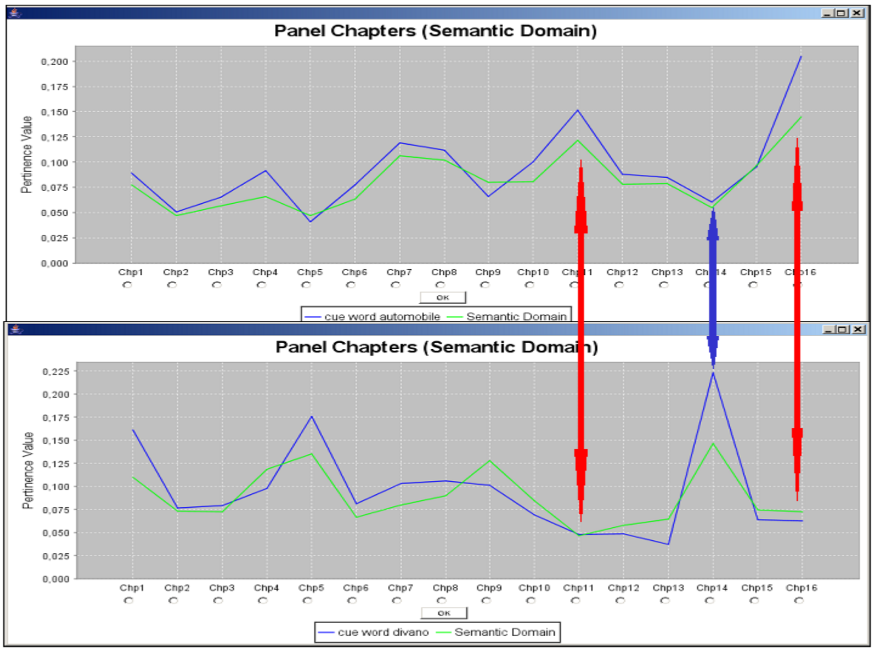


Fig. 2. The plots of two semantic domains, *automobile* (upper panel) and *noia* (lower panel), across the chapters of "*Gli Indifferenti*"

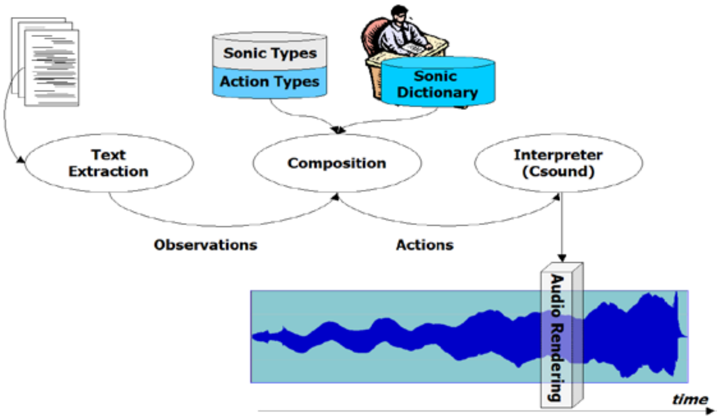


Fig. 3. An overall view of the sonification process

Sonic Types In music information processing two major representation forms exist: symbolic and audio. A third type is also extensively used although a comprehensive standardization has never reached a reasonable maturity. This form determines parametric definitions used to constraint the behavior of music generator devices, such oscillators in the FM synthesis [13]. In the sonification system implemented for this study an independent Sonic Type for each of the three kinds of representation has been defined. For example, a ST for audio samples defines information for the audio format (e.g. MP3) or the cross-reference to its corresponding physical storage file, while a ST for MIDI sequences supports descriptions such as length or tempo as well as defaults expressing aspects such as instrument class or musical scale. An instance of a Sonic Type is declared in the Sonic Dictionary of a composition, and is manipulated by a finite set of parametrized actions. Two Sonic Types may represent the same class of musical information (e.g. MIDI) but may be characterized by different parameters and actions. A Sonic Type is the (Java) object responsible to represent a musical object and implement all the actions decided during composition. They respond to the decoding process fed with literary information from the text. According to this strict OO view, newer Sonic Types can be defined as specialization of existing types and refinement of some properties and actions of the latter ones. This abstraction allows to detach the text analysis and decoding subprocesses from the design of new sets of actions, aiming to increase flexibility towards more complex object types and meaningful musical criteria.

The Sonic Dictionary The Sonic Dictionary includes all the information about a specific composition. Its design is totally under the responsibility of the musician, where he declares which music material is necessary for the targeted composition and which defaults characterize the individual (typed) musical objects. It works as a constraint on the stochastic composition process. A typical parameter that is crucial to the targeted opera is the duration of the entire song. By specific choices in this phase the musician can characterize important timbral elements of a song as well as constraining the relative freedom of the stochastic composition. A relevant choice in the Dictionary is the assignment of specific (sets of) concepts to individual musical objects (i.e. objects of some Sonic Types): in this way the text extraction process regarding a given concept c is imposed to influence a specific music object o, and its actions to manipulate o correspondingly. This mapping can be thus used to create particular effects on individual layers of a composition, where timbral properties can be harmonized with the particular bias (positive/negative, happy/sad) of a concept. If an existing Sonic Type instance is declared in a Sonic Dictionary but is not mapped into any concept, it will not be used in the song. This allows to create large libraries of musical objects and apply them only in specific cases. Notice that a simple declarative language has been defined for Sonic Type instances so that they are directly realized into Java objects at runtime. Sonic Types here guarantee that the proper actions will be used and declarations alone suffice to determine the proper computational behavior of individual musical instances.

Actions Types A primitive action is a manipulation of a musical instance as a change over one ore more of its fundamental properties. Actions are parametrized in order, for example, to determine different strengths of its consequences on a musical instance. Actions are applied at given predetermined time points, called *ticks*, that form a segmentation of the duration of a song. Notice that ticks correspond to units of the source text (e.g. chapters or paragraphs) so that at every tick the observation of a concept c in the text is made available. At a given tick t an action (decided during the decoding process) is applied with a given *strength*. An action may persist for a number of ticks (or time segments), according to a second parameter called *scope*. This value determine the time interval during which the action may persist. Also strength and scope are made dependent on the input sequences of observations from the text. In case a proportional law is implied, the higher the relevance of a given concept in a portion of the text, the higher will be the strength in the corresponding tick.

Text units and musical events Values of relevance for a concept c in the text are computed according to equations 3 and 4 as observable over units, such as chapters or paragraphs. Any ticks always corresponds to an observation from which an action can be triggered and parametrized. Larger units, such as chapters, give rise to sequences of macro-units, called *macro-sequences*. If paragraphs are used in Eq. 3, sequences of micro-units are obtained, called *micro-sequences*. In order to guarantee a good granularity to the sonification process, ticks are usually made correspond to micro-sequences. Notice how macro-sequences represent larger text structures and may be used to model macro-properties of a song, such as the distinctions between *intro* and *chorus-like* phases of a composition. Notice that a null action is defined and even when highly granular actions are possible (for shorter tick intervals), more static behaviors can be obtained from specific null action sequences, that leave longer song segments unaltered. The relationship between time ticks and actions is represented in Fig. 4. In the figure, at every tick t, every Sonic Type TS_k corresponds to actions among those enabled by the Sonic Dictionary to be applied to the corresponding instance.

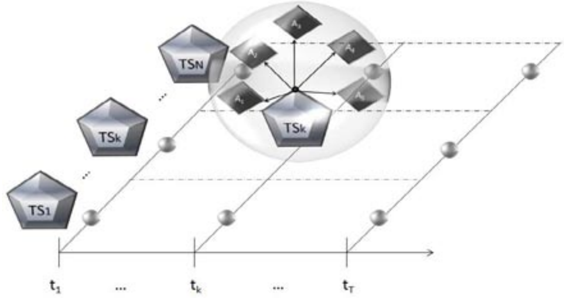


Fig. 4. Discrete representation of Sonic Types and the corresponding Actions

Given the t-th tick every Sonic Type instance is modified according to the corresponding action, a given scope and strength. The action, but also the strength and scope parameters, depend on the t-th value observed in the input observation sequence. Actions are selected as a side effect of an HMM-based decoding process, as described in III-A.

The FM Generator Sonic Type As a first full implementation of a Sonic Type is given by a generator based on the frequency modulation synthesis (or *FM synthesis*, [14]). It is a form of audio synthesizer where the timbre of a simple waveform is changed by frequency modulating it with a modulating frequency that is also in the audio range. This results in a complex waveform with a different-sounding tone. In this case the manipulation has more degrees of freedom and more interesting timbral effects can be obtained through the combined perturbation of simple sinusoids. In this work a configuration based on four individual and independent oscillators is employed, thus forming sets of FM operators. An operator can be controlled through the two simple parameters of *frequency* and *amplitude*. The different configurations allowed by four FM operators are shown in Figure 5.

In 5.*b*, for example, frequency modulation is not employed, while a single carrier with a repeated modulating wave is shown in 5.*c*. Notice how combinations of these cases are possible as shown in 5.*a*, 5.*f*, 5.*g*, 5.*h*, 5.*i*. A FM generator instance is defined in the Sonic Dictionary with a defalt configuration: however, its initial configuration can be made dependent on the initial observations as received from the text (at the t-th tick for $t = 0$). Although the number of parameters in an FM generator is relatively small, it offers a rich timbral variety that made it a very powerful synthesis device. Typical applicable actions are related to the treatment of frequency and amplitude, so that the following six parametric actions, described in Table I are used: Amp_Up(m_t), Amp_Down(m_t), Amp_Hold(m_t), Freq_Up(m_t), Freq_Down(m_t), Freq_Hold(m_t).

Action Type Name	Description
Amp_Up(m_t)	increases amplitude
Amp_Down(m_t)	decrease amplitude
Amp_Hold(m_t)	no modification to amplitude
Freq_Up(m_t)	increase frequency
Freq_Down(m_t)	decrease frequency
Freq_Hold(m_t)	no modification to frequency

TABLE I
ACTIONS RELATED TO THE FM GENERATOR SONIC TYPE

A. *Actions as HMM states*

The discrete and syntagmatic nature of a text is used in this work to translate it into a sonification process. The mapping between text units and time ticks can be refined to make the action selected at time t depend (1) on the action (state) decided at time tick $t-1$ and on the observed relevance of a concept observed at time t, i.e. Equation 3. As actions are states in the process and relevances correspond to signals, the process of generating action sequences can be seen as a Markov process. In particular the hidden states characterize it as an Hidden Markov process. In Figure 6, three

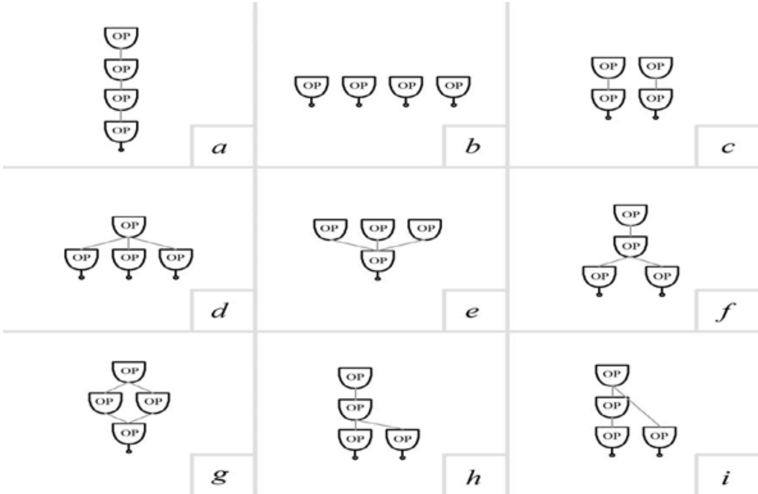


Fig. 5. Possible configurations from 4 FM operators

Action(/States) are shown with their corresponding transition probabilities, p_{12}, p_{23}, p_{23}, p_{31}, p_{32}.

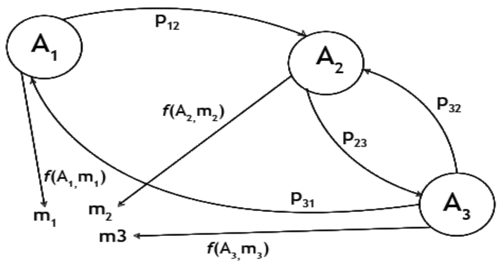


Fig. 6. An HMM for sonification with 3 State/Actions

As we use micro-sequences of individual tick chains at time t, there is a corresponding micro-unit from the text observed with a relevance value of m_t. The probability that an Action/State is emitting m_t at time t is computed as a $f(A_x, m_t)$, as it depends on both information. In the example of Figure 6, the probability that the Action/state A_1 outputs a symbol m_1 is $f(A_1, m_1)$. Correspondingly , $f(A_1, m_2)$ is the probability that A_1 outputs m_2. The application of the well known Viterbi algorithm, given an initial distribution of probabilities among the initial states allow to efficiently compute the most likely state chains, i.e. the sets of events that correspond to the targeted stochastic composition. The audio rendering of the events has been obtained in CSound using the fairly easy notions of file orchestra and score according to the templates for FM synthesis instruments.

IV. A case study: Sonification of "*Gli Indifferenti*"

In order to validate and experiment the above defined model for narrative analysis we made a quantitative study of the novel *Gli Indifferenti* by Alberto Moravia [12]. The book (corpus C) is made of about 16 chapters and about 91,059 words (tokens). The number of different words in the novel is 3,273. Additionally, we created the extended corpus C_E by including critical reviews up to a total size of 13,041 tokens with 3,920 different words. Individual pseudo documents have been created from the opera based on paragraphs: each pseudo document consists of a single paragraph in the opera. We found about 1,854 total paragraphs and 116 paragraphs per chapter on average.

Different weighting schemes can be adopted for the assignment of initial values to the term-document matrices that triggers LSA. The score adopted in all the experiment discussed in here is the simple lexical frequency tf_{ij} of words w_i in the pseudo-documents t_j. The dimensions used by the

LSA on both corpora have been limited to (the first) 100 dimensions (i.e. principal components).

The generation of the semantic domains has been obtained through the notion of distance in the LSA_E space. In the different runs a threshold value (i.e. τ in Eq. 4) of 0.5 has been applied: in general about 35 different lemmas are obtained in the lexicons L_c, of which about 10 are nouns.

The system can be activated with every abstract concept as originating *cue word*. Some examples of semantic domains derived from meaningful cue words (c) are the following:

- c=*noia*. L_{noia} = { *esistenza, atto, avventura, fatalita', falsita', familiare, ragazza, abitudini, ...* }
- c=*indifferenza*. $L_{indifferenza}$ = { *vita, noia, scena, prova, vero, volonta', esistenza, proposito, ambiente, mancanza, incapacita', vanita'*}
- c=*Carla*: L_{Carla} = { *osservare, baciarono, torpore* }
- c=*Leo*: L_{Leo} = {: *suonare, ingiunse, stiro', cammina, fastidio, signor*}

The resulting lexical descriptions are very interesting as a number of semantic phenomena are captured in a fully automatic way. First, words strongly correlated with the cue word are derived (e.g. *noia/boredom* vs. - *abitudini/habits*, *esistenza/existence*). Second, correlation at the level of the typical plot of the novel are also obtained, like the *noia-falsita'/falsity* pair. Notice here that the notion of *falsity* is a strong connotation of the typical middle class family described in the novel: it is a sort of originating state of the boredom itself.

The musician intervention here aims at defining the background musical material to characterize some timbral and dynamics aspects of the target song, and, then, mapping these latter to the availale concepts described in terms of semantic domains, e.g. *noia*. Examples of the subsets of semantic domains to which a single FM generator instance is mapped are reported in Table II, where the cue words c, and the subset of the lexicons L_c for different semantic domains are shown. Notice how the same cue word (i.e. semantic domain) is used to build different word sets as triggers for an individual musical object, such as one FM Generator, as row 3 and 4 in Table II show.

Cue word	Subset of the semantic domain
automobile	automobile, corsa, persone, pioggia
divano	budoir, gambe, desiderio
divano	gambe, ginocchia, carne, libidine
Leo	fastidio, ingiunse
noia	fatalita', falsita', familiare

TABLE II
CUE CONCEPTS AND LEXICONS MAPPED INTO AN FM GENERATOR

In Figure 7 the wave form derived from a composition based on the cue word *noia* is shown, while 8 is the result of the composition based on the cue word *divano*: although in a paper we cannot appreciate the musical differences the two plots indicate a completely different behavior that seems to express different semantic implications for the two (quite unrelated) semantic domains.

V. CONCLUSION

A specific use of linguistic creativity is the automatic musical composition inspired by semantic analysis of texts. This task consists in the translation of complex concepts, as they manifest in a literary text, into a musical piece. In this paper methodological and technological aspects of this process have been discussed. In particular, a markovian composer acting over streams of (literary) concepts extracted through latent semantic analysis over the target text has been defined. Its application to the composition of short electronic pieces based on FM synthesis has been discussed through the definition of an object-oriented software platform. Its application to the study of a novel, "Gli Indifferenti" by Alberto Moravia, has been presented. While the AI approach here used is tailored to a reatively simple case of sonification, the proposed system can be seen as a basis for more complex composition models. It allows novel approaches with interesting aesthetic perspectives opened by the combination of semantic language modeling and automatic composition, towards sophisticated forms of expressive musical performance [15].

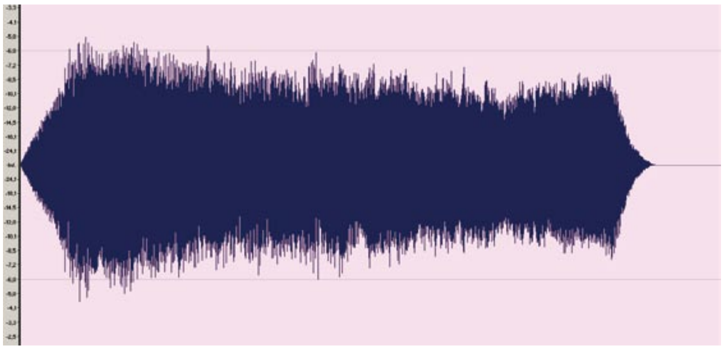


Fig. 7. The sound wave form of the ”*Noia*” composition

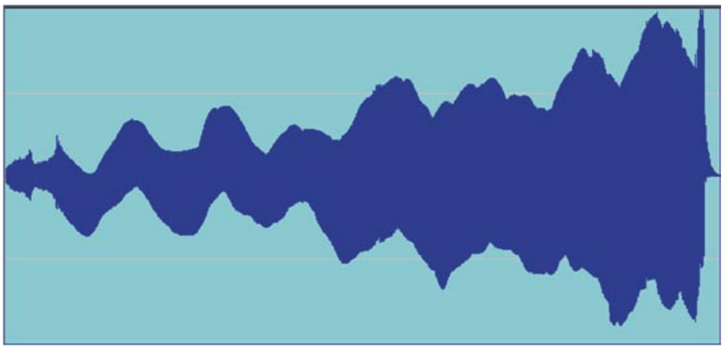


Fig. 8. The sound wave form of the ”*Divano*” composition

References

[1] I. Xenakis, *Formalized Music: Thought and Mathematics in Composition*, 2^{nd} ed. Pendragon, 2001.

[2] H. S. C. Manning, *Foundations of statistical natural language processing*. MIT Press, 1999.

[3] H. L. and Isaacson, *Musical Composition with a High-Speed Digital Computer*, S. M. Schwanauer and D. A. Levitt, Eds. Cambridge, Mass.: MIT Press, 1993.

[4] L. Rabiner, “A tutorial on hidden markov models and selected applications in speech recognition,” in *Proceedings of IEEE*, 1989.

[5] M. Sahlgren, *The Word-Space Model*, Department of Linguistics, Stockholm University, 2006.

[6] T. Landauer and S. Dumais, “A solution to plato’s problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge,” *Psychological Review*, vol. 104, pp. 211–240, 1997.

[7] T. Hofmann, “Probabilistic latent semantic analysis,” in *Proc. of Uncertainty in Artificial Intelligence, UAI’99*, Stockholm, 1999. [Online]. Available: citeseer.ist.psu.edu/hofmann99probabilistic.html

[8] T. J, V. de Silva, and J. C. Langford, “A global geometric framework for nonlinear dimensionality reduction,” *Science*, vol. 290, pp. 2319–2323, 2000.

[9] G. W. Furnas, S. Deerwester, S. T. Dumais, T. K. Landauer, R. A. Harshman, L. A. Streeter, and K. E. Lochbaum, “Information retrieval using a singular value decomposition model of latent semantic structure,” in *Proc. of SIGIR ’88*, New York, USA, 1988.

[10] A. M. Gliozzo, *Semantic Domains in Computational Linguistics*. University of Trento, 2005.

[11] G. Vigliocco and D. Vinson, “Semantic representation,” in *Handbook of Psycholinguistics*, G. Gaskell, Ed. Oxford: Oxford University Press, 2005.

[12] A. Moravia, *Gli Indifferenti*. Bompiani, 1929.

[13] C. Roads, *The Computer Music Tutorial*. M.I.T. Press, 1996.

[14] R. Boulanger, *The CSound Book, perspectives in software synthesis, sound design, signal processing and programming*. M.I.T. Press, 2000.

[15] J. L. A. Ramon Lopez de Mantaras, “Ai and music: From composition to expressive performance,” *AI Magazine*, vol. 23, no. 3, pp. 43–58, 2002.

Cultural Heritage and the Web: a Perspective from Archaeology

Glauco Mantegari
Complex Systems and Artificial Intelligence Research Centre (CSAI)
QUA_SI Doctoral and Advanced Research Program
Department of Informatics, Systems and Communication
University of Milan Bicocca, Italy
glauco.mantegari@disco.unimib.it

Abstract

The paper provides an overview of some key elements in the scenario of Web applications in Cultural Heritage. Archaeology is taken here as a discipline that offers several ideal characteristics for Web applications, thanks to its very composite nature which comprises highly interrelated and often incomplete information of different kinds. In particular, developments in the field of geographic Information on the Web and in the field of Semantic Web are introduced and discussed in light of the Cultural Heritage and Archaeology scenario. In fact, these topics are receiving increasing attention and may offer great opportunities for the representation, management, retrieval and dissemination of archaeological information, both for specialists and for the general public. The "Ancient Milan" project is then briefly described as a testbed for the application of these new approaches and technologies.

1. Introduction

Web technologies represent a great opportunity for Cultural Heritage Management (CHM). Over the years the diffusion of the Internet determined the birth of a plethora of projects dealing with the dissemination of Cultural Heritage information, and the objectives and results of these experiences vary a lot. However, scores of projects are increasingly dedicated to this topic along with the number of scientific events dealing with it. So much interest comes from the perception — rather than the recognition — that today's technologies offer new perspectives and possibly effective solutions for the entire spectrum of CHM activities, which range from scientific research to heritage valorization and public fruition.

Within the vast field of Cultural Heritage, Archaeology represents the discipline where computer applications have been experimented and used the most: just to mention an example, an international conference, "Computer Applications and quantitative methods in Archaeology" (CAA)[1], has been held on this topic every year since 1973. In fact, Archaeology offers a very rich set of methodologies, contexts and problems that it is somehow impossible to identify a specific case where computers do not find a possible application. However, if a large number of experimentations have been undertaken, their results often appeared much below the initial expectations, and this consideration is particularly valid in the case of Web applications. While specific discussions about the causes of this situation can be found in the literature (starting from CAA proceedings), a few elements are mentioned here in order to introduce the scenario this paper considers.

One key element is represented by a kind of naïve approach to Web technologies, which have often been considered as something between a fun toy and a to do activity; in fact, it seems that a large number of projects have been undertaken only to give archaeological research a "modern" touch. The deep investigation and understanding of the potentials and, most of all, the limits of these technologies have been left behind, in favor of the acquisition of the basic competences that are required in order to create "the website", thus producing a distorted interdisciplinary perspective. From another point of view, the academic tradition itself has frustrated a more fruitful exploitation of these technologies: electronic publishing has been considered much less valuable than the venerable paper publishing, and protectionism on data has impeded Web applications to show their potentials.

Of course, over the years exceptions have emerged, and interesting initiatives were undertaken for a more accurate and effective use of the technologies and the Web within Cultural Heritage and Archaeology, such as, among the others, Minerva[2] and DigiCULT[3]. Moreover, it has to be ac-

[1] http://www.leidenuniv.nl/caa
[2] http://www.minervaeurope.org
[3] http://www.digicult.info

knowledged that the increasing complexity of Web technologies both in conceptual and in technical terms makes the current scenario full of potentialities but also full of new challenges. This consideration is particularly true if we examine two of the most prominent areas in the current scenario which may offer innovative opportunities for Web applications in Cultural Heritage and in Archaeology: the Geospatial Web and the Semantic Web.

The paper is organized as follows: sections 2 and 3 provide some insights in the Geospatial Web and the Semantic Web from the point of view of their application to the archaeological research; section 4 illustrates the main elements that characterize a joint project that was recently undertaken in order to experiment innovative Web applications in the context of the ancient city of Milan; section 5 draws some concluding remarks.

2. From Internet GIS to the Geospatial Web: exploiting space in the archaeological Web

Space in its broadest sense (from geometry and location to topology) represents a fundamental property of everything: it is estimated that almost 80% of all data are geospatial data [6, p. 3]. In Archaeology, space was acknowledged as a fundamental dimension for the understanding of the past since the origins of the discipline; however, it was only from the late '60s and the '70s that its analysis in the archaeological contexts received specific attention [18, 9]. Moreover, starting from the '90s, developments in spatial technologies made GIS a successful and popular application also within the community of archaeologists: in fact, these tools proved their effectiveness in scientific research [31, 10] and dominated the scenario of computer applications in Archaeology over a decade.

2.1 Internet GIS and archaeological data

Recently the representation and management of geospatial data (i.e. the ones dealing with the Earth's surface and near–surface) has been a major topic in the scenario of Web applications. The massive diffusion of the Internet has offered fertile ground for experimenting innovative approaches, and since the beginning of the '90s much effort has been spent in the creation of the infrastructure for distributed GISystems. Internet GIS [25], alternatively known as "WebGIS" or "On-line GIS", constitutes the result of these efforts and its success has been remarkably rapid: the Xerox PARC Map Viewer [26], which can be considered the "grandfather" of Internet GIS, was receiving over 25.000 maps image request per week within the first 9 months of its release, and was soon soon followed by several other pioneering projects, such as the National Atlas Information Service by Natural Resources of Canada (1994), the Alexandria Digital Library (1994), and GRASSLinks by the University of California at Berkeley (1995). This success was mostly due to the advantages offered by Internet GIS in providing effective solutions to several issues that are traditionally related to desktop GIS, such as the costs of data collection and updating (see e.g. [21]), which mainly derive from redundant geospatial data collection (see e.g. [22]). Moreover, a number of encouraging possibilities soon emerged, such as the democratization, open accessibility, and effective dissemination of spatial data, and stakeholder participation enhancement, just to mention a few [14].

On the other hand, several new issues soon had to be faced, mostly because the inherently complex nature of a networked environment made geospatial data management more difficult than in traditional desktop systems [24]. Moreover, other problems emerged beyond the technological ones, such as reliability, accuracy and copyright of geospatial data, and privacy issues connected with the possibility of virtually visualizing nearly every part of the Earth's surface in great detail.

This debate did not find a particular echo into the community of Cultural Heritage and Archaeology professionals, which began using Internet GIS technologies in a plethora of different projects, from applications concerning single excavation projects and limited territorial areas to large scale interactive mapping covering larger areas (e.g. "POBASyN"[4]), one or more countries (e.g. "FastiOnLine"[5]), a continent or even the Globe (e.g. the mapping of UNESCO's heritage sites[6]).

In these experiences Internet GIS had been used mostly for data management, with objectives varying from the sharing of information to collaborative editing of data within a research group. This situation was similar to the one that occurred during the first use of GIS in Archaeology, even if it has to be admitted that Internet GIS technologies were not mature enough to provide more than simple data management and display. However, the scenario seems to be changing, thanks to the improvements that made geospatial technologies evolve into the bigger scenario of the Geospatial Web.

2.2 Beyond Data Management: the Geospatial Web

A decade after the release of the first Internet GISystems, a new phase began, in which the technological developments and the effects of geospatial data usage became more and more intertwined. The "Geospatial Web", or "GeoWeb" [27], represents the result of this evolution, and its

[4] http://www.archeoserver.it/pobasyn

[5] http://www.fastionline.org

[6] http://whc.unesco.org/en/list

birthday can be symbolically set at February 8, 2005, the day Google launched its "Maps" platform[7] (which was very soon followed by "Earth"[8], in June), which was destined to boost the capabilities of Internet GIS. At first, the most distinguishing elements of Google products were the superior map and image quality and the surprisingly rich and easy possibilities of interaction with geospatial data. However, the really innovative characteristics of the new platforms were beneath the surface. In fact, for the first time simple tools for accurate geospatial data creation were provided, which did not require almost any expertise in geospatial technologies; moreover, an official API was released in order to help even low skilled programmers in personalizing the map contents and behavior. For the first time in history the creation of professional quality maps overcame the boundaries of professional cartographers: from the users' point of view maps suddenly changed from a "read only" to a "read-write" medium, as suggested by [15], allowing almost everyone to publish his own geospatial content.

This user generated content (UGC), a peculiarity of the so–called "Web 2.0", made geospatial information become a central element of the new Web scenario to the point that big specific events where organized (such as the "Where 2.0" conference by O'Reilly) and even a new term was proposed (neogeography [30]) in order to suggest the birth of new modalities of geographical information production, management and sharing. Even if geographers and GI-Scientists rightly stress that these amateur observations are very different from the rigorous methodologies of Geography as a scientific discipline (see e.g. [16]), it has to be acknowledged that the Geospatial Web represents a milestone in geography and GIS history, and it has to be considered as a central element of today's GIScience research.

Experiences in the field of Geospatial Web applications in Archaeology are just beginning. At a basic level, Google Earth and its counterparts (e.g. Microsoft Virtual Earth[9]) have been evaluated by researchers for the identification of archaeological sites, thanks to the free high resolution and constantly updated satellite images provided [23]; and of course, the mapping and sharing of archaeological heritage information have continued in the new scenario. Some novel elements are the use of standard Geospatial Web platforms, rather than custom ones, and the representation and sharing of data, which is performed through standard formats and protocols (e.g. GML, KML, WMS, and WFS). In fact, the developments in Internet GIS and the advent of the GeoWeb coupled well with the spectacular diffusion of the Open Source paradigm and this marriage gave birth to the creation of open and standard formats and protocols, mostly thanks to the activity of the "Open Geospatial Consortium" [10] and ISO/TC211[11].

However, the potential applications of the GeoWeb in Cultural Heritage and Archaeology are a lot more and go far beyond simple data management. While an exhaustive overview of the scenario is impossible here, some points of important general interest can be summarized:

- Visualization: importing and overlaying personal georeferenced raster images or vector shapes into different platforms, or even georeferencing and drawing them directly is no longer a problem, thanks to the standardization of data formats and to the large number of available tools. Moreover, impressively efficient virtual globes (i.e. 3D representation of the Earth, such as Google Earth and similar) are available which offer the possibility to incorporate and navigate through 3D rendered models of buildings, excavations, etc. (see e.g. Google 3d Wharehouse[12]). This is a ready–to-use, zero costs and effective alternative to several highly expensive custom applications which have been developed with the same intentions over the last few years.

- Analysis: even if several constraints are currently present that make the geospatial and geostatistical analysis hard to perform in a Web environment (e.g. hardware resources, possible instability of the connection, etc.) they will probably disappear in a relatively short period of time. Moreover, some WebGIS backends (such as PostgreSQL–PostGIS[13]) already offer sophisticated geospatial analysis functionalities

- Multimedia: almost any kind of media format can be hyperlinked to georeferenced data, thanks to the nature of the Web, thus overcoming the traditional and well known difficulty of desktop GIS in managing multimedia. Moreover, relevant content can be retrieved automatically and "mashed–up" via Web services from other trusted repositories and applications.

- Social participation: non–professionals can participate in the heritage management process in a variety of way, e.g. indicating their observations on a map.

These elements are very interesting (even if potential issues related to them have to be taken into consideration, and cannot be discussed here), but what may really boost the GeoWeb impact will probably be the adoption of semantic techniques for data integration and retrieval. This particular scenario is not only acknowledged as one of the most promising ones in the overall GeoWeb development ([27]), but it represents also a great opportunity for Cultural

[7] http://www.maps.google.com
[8] http://www.earth.google.com
[9] http://www.microsoft.com/virtualearth
[10] http://www.opengeospatial.org
[11] http://www.isotc211.org
[12] http://sketchup.google.com/3dwarehouse
[13] http://postgis.refractions.net

Heritage and Archaeology where semantic models and techniques – starting from the Semantic Web – are increasingly investigated, as will be very briefly discussed in the next session.

3. The Semantic Web: giving meaning to the archaeological Web

Archaeology is a composite discipline in which a large number of specialized studies converge. This characteristic determines a huge variety and heterogeneity of interconnected information, which is often difficult to effectively dominate and disseminate not only to the general public but also to professionals. Moreover, information about the past is often incomplete, thus posing several problems in terms of its evaluation and interpretation. In this framework, the representation and management of archaeological information by means of computer based techniques and tools are very challenging. The Web in theory offers a good approach to represent, retrieve and display inter-related information (e.g. via hyperlinks), but in reality the situation is very different, as is experienced everyday by Internet users. To date the Semantic Web [2] is widely recognized as the most articulated proposed solution for a variety of problems concerning the "traditional" Web: a great deal of research has been done on it, but the original vision still remains largely unrealized [29].

The characteristics of the Cultural Heritage and Archaeology domain potentially offer several challenges for Semantic Web research, both from a theoretical and from a technological point of view; and conversely the Semantic Web promises innovative approaches for the issues related to the representation and retrieval of Cultural Heritage and Archaeology information in electronic forms.

In fact, these potentialities have been recognized, and interdisciplinary research on these topics is increasing, as the number of specific events and publications show (see e.g. [1]).

Several projects based on the definition of domain ontologies and on their usage for effective information representation, sharing and retrieval have been proposed and experimented. Most part of the efforts have concentrated on the problem of the integration of metadata from different resources, with the aim of granting a uniform access to them. The definition of a standard Conceptual Reference Model, the CIDOC CRM [13], represents the most important result of research activity in this area; its development lasted for over 10 years within the activities of the Comité International pour la Documentation des Musées of the International Council of Museums (CIDOC-ICOM[14]). Initially defined as a knowledge representation model to achieve semantic interoperability of museum data, CIDOC CRM currently provides an extensible ontology for Cultural Heritage and museum documentation, which as of May 2008 [11] is made up of nearly 100 classes and more than 100 logical properties linking them. In december 2006 the model has been accepted as the international ISO 21127:2006 standard, thus becoming a mandatory reference for all the projects facing the problems of semantic information integration in the domain of Cultural Heritage.

Several applications testify the acceptance and diffusion of the CIDOC CRM, some of which are listed in the project web pages[15]. Moreover, research initiatives have been undertaken in order to increase its effectiveness in a number of situations: mappings of other standards into the CIDOC CRM Core metadata element set have been proposed, from international standards (e.g. Dublin Core) to national ones (e.g. the national standards for archaeological excavation data recording); harmonization projects have been defined and specific extensions of CIDOC CRM for the Archaeological domain have been created (such as the extension by English Heritage in the context of the STAR project[16]), and are being tested [3].

On the other hand such a general model as the CIDOC CRM while offering a superficial homogeneity may mask low-level conceptualizations, especially when mapping legacy datasets. Research in the direction of semi-automated mapping tools to CIDOC CRM is done under the AMA project [17] and more articulated studies concerning the theoretical, methodological and practical issues of more dynamic inter-alignments of different ontologies is done under other projects, notably the TRANSLATION framework project[20].

The scenario is vibrant: the basic building blocks for Semantic Web applications in Archaeology are available, and significant improvements are expected in the next few years. What is missing is essentially a more extensive evaluation of these technologies in a sufficient number of concrete and articulated case-studies and experimentations. One impediment is represented by the objective difficult and money/time consuming activities of mapping legacy datasets to ontological schemas; in fact, this activity currently often requires highly technical skills that may be out of budget for Cultural Heritage and Archaeology institutions. More in general, semantic approaches require the adoption of a new mindset that may be hard to achieve, especially for archaeologists who worked with traditional models (e.g. entity-relation) and technologies for years.

From another perspective, semantic approaches are increasingly investigated in the context of GIS and, especially, in the GeoWeb (see e.g. several chapters in [27]). The Se-

[14] http://cidoc.mediahost.org

[15] http://cidoc.ics.forth.gr

[16] http://hypermedia.research.glam.ac.uk/kos/CRM

[17] see http://www.epoch-net.org

mantic Geospatial Web is given the capability to add space and time dimensions to the Semantic Web and as such, it is easy to understand how much its relevance could be. Moreover, given the crucial importance of space within Cultural Heritage and Archaeology, the coupling of these perspectives with the ones that were briefly described in this section represents an exciting yet difficult challenge. Nevertheless, it may offer significant improvements not only for a more effective integration, retrieval, dissemination (and more in general "use") of archaeological information, but also for the definition of new methodological approaches, beyond the merely technological issues that were often privileged in the last few years. The "Ancient Milan" project, which will be introduced in the next section, aims to become a testbed for these scenario.

4. An application Scenario: the "Ancient Milan" Project

Milan, one of the most important cities in Italy, has an important history dating back to the 5th century B.C., when Insubrian Celts settled in the area for the first time. From that period on, the city extended progressively, especially under the Roman domination. In the Tetrarchy period, and precisely between 286 and 310 A.D. Milan became the seat of Maximinian imperial residence, thus acquiring great importance in the Roman empire. Constantine emanated his edict granting freedom of worship to Christians in the city in 313 A.D.; while, from 374 to 397 A.D. the famous Bishop Ambrose ruled spiritual and political life and gave the city a Christian physiognomy. However, during the 5th century A.D. the ancient city started its fall into decay: the imperial seat was transferred to Ravenna (402 A.D.) after the invasion of the Visigoths, and shortly thereafter later the city was sacked by Attila (452 A.D.) and finally destroyed by the Goth Uraia in 538-539 A.D.

Such a great history, which lasted for a millennium, makes Milan's archaeological heritage particularly significative [7]. Researches are continuously conducted in the city in order to better reconstruct and understand the setting of the ancient city. It is an everyday experience that during public works in the historical centre, remnants of ancient buildings come to light. If this richness represents a fascinating element for the occasional visitor to a very modern city, it also raises a number of problems concerning the activities of heritage study, from research to valorization.

"Ancient Milan" is born from a project funded by Regione Lombardia and concerning the study and experimentation of innovative archaeometrical techniques for the chronological attribution and certification of materials coming from excavations conducted in the city; thise project funded by Regione Lombardia. The leading partner of the project is the Milano Bicocca Centre for Dating (CUDaM[18]) at the University of Milan–Bicocca, directed by Prof. M. Martini, which works in collaboration with the Archaeological Museum in Milan. The scope of the project was recently extended in order to experiment innovative IT supports both for the dissemination of the project results and, more in general, for the "discovery" of the Archaeology of Milan by professionals as well as by the general public (see section 4.1). The IT partner is the Complex Systems and Artificial Intelligence Research Centre at the University of Milan Bicocca[19], which is involved in research on IT models and technologies for the Cultural Heritage since its foundation in 2007.

The final objective of the entire project is the definition and development of technologies supporting the discovery, valorization and fruition of the archaeological heritage within the area of the city.

4.1. Defining the requirements

Web technologies can represent a key element in the scenario discussed here. In fact, most of the monuments are badly preserved, and they are often incorporated into later buildings due to urban development. As a consequence, it is not only difficult to visit the remnants of the excavated structures, but it is also extremely difficult to perceive their significance and relationships within the context of the ancient city.

The very initial definition of the requirements took three main objectives into consideration:

- to support an archaeologically aware visit of the heritage in the urban area, which means that it is not sufficient to provide more or less sophisticated virtual replicas of the monuments;
- to timely and extensively communicate the results of the latest researches, both for the specialist and for the general public;
- to provide easy and selective access to information and specific documentation coherently with the user interests and profile.

In order to reach these objectives it is important to define the basic dimensions which have to be taken into consideration when representing, displaying and retrieving data. In particular, the dimensions currently defined can be grouped into three main classes of properties:

- spatial properties: the location, extent and spatial relationships each structure had and currently has within the context of the city;

[18] http://cudam.mater.unimib.it
[19] http://www.csai.disco.unimib.it

- temporal properties: the phases that characterized each structure during time, including eventual changes in its functional use and its post–abandonment transformations;
- qualitative properties: describing, for example, differences in functions, materials, styles, etc.

Moreover, it is important to stress that the project will not embrace only a traditional historical or art–historical perspective, but it will extend its scope to the processes of discovery, study and preservation of the monuments. In our opinion this information is very important, mostly because it can be considered as part of the history of each archaeological item and as such it has to be properly evaluated and understood. Suggestions for accomplishing this task come from the advancements in the study and representation of processes by means of formal models, such as, for instance, Petri Nets and business processes modeling languages.

Given these general requirements, It is easy to understand how the GeoWeb and the Semantic Web may offer several opportunities to do this. And, of course, a number of challenges emerge from the scenario, starting from the problem of combining different technologies into an integrated system. In fact, single solutions currently exist that may satisfy each requirement: GeoWeb technologies offer a variety of approaches for the representation and management of the spatial components and standard semantically-enabled models are available for the Cultural Heritage domain (such as the previously introduced CIDOC). However, a few experiences in Cultural Heritage have currently considered a composite scenario like the one we propose. The integration of existent, standard and robust Web services and the possibility to mashup contents are often considered to be a promising approach in this sense (CITARE), but their real effectiveness has still to be evaluated beyond the technological results.

On the other hand, several aspects still constitute open research issues within specific disciplines, such as the problem of the representation of time in GIS (see e.g. [8]), not to mention the effective definition of semantic retrieval facilities.

The project in its entirety is ambitious and as such has to be considered in the long term perspective. However, a prototype is under development and is being tested in order to progressively approach all the issues that characterize the scenario. This prototype makes use of an existing framework, the MIT "Exhibit"[20] [19], which is an interesting ready–to–use solution combining a set of Javascript–based tools that simplify navigation in complex datasets . Among other facilities, Exhibit offers the possibility to integrate and to coordinate the behavior of an interactive map with an interactive timeline and a customizable facet browser. Navigation through facets seems to be currently a good solution for the effective filtering of information, especially for all the users who do not precisely know what to search for. Experiences conducted in recent years have shown that this kind of "exploratory" research may be very useful. In the domain of Cultural Heritage sophisticated forms of faceted browsing are under development and experimentation on large scales, such as in the case of the MultimediaN E-Culture Project [17, 28].

The prototype represents a starting point for the evaluation and discussion of the interaction modalities with the partners of the "Ancient Milan" project. Of course, several limitations come from the architecture of the framework, which cannot be discussed here in details. Just to mention an example, the display of complex cartographic features is impossible, but a lot of more sophisticated alternatives are available, comprising a set of technologies (i.e PostGIS, UNM Map Server and Open Layers) that were successfully experimented in another project, which we are currently involved in[21].

Moreover, the prototype totally lacks the semantic support that is required in order to enrich its browsing capabilities. It is our intent to experiment with the combination of faceted browsing with domain ontologies, and in particular with the standard CIDOC Conceptual Reference Model. This approach represents an innovative and open research area within the interdisciplinary field of Computer Science Applications to Archaeology, and only a few experiences in this direction were recently undertaken in the broader area of Cultural Heritage, such as the above mentioned MultimediaN E-Culture Project. Experiences in the field of ontological annotation of Cultural Heritage contents and semantic retrieval have been made in different projects our Research Centre is involved in, such as the ones described in [5, 4], which are specifically related to Archaeology. A knowledge-based approach may also be an effective solution for managing the behavior of the system according to different user profiles. Related work in our department concerning multi-view and multimodal interaction based on semantic profiling has recently taken into consideration an archaeological scenario [12].

5. Concluding remarks

This paper provided an overview on Web applications to Cultural Heritage and Archaeology, focusing specifically on two emergent and innovative perspectives: the Geospatial Web and the Semantic Web. These technologies have great relevance for the effective structuring, management, retrieval and dissemination of archaeological data in elec-

[20] http://simile.mit.edu/exhibit

[21] http://www.archeoserver.it/pobasyn

tronic forms. In fact, space represents a fundamental dimension of almost every archaeological item, and as such, it has to receive specific attention; in addition, the structure of the archaeological information is very articulated and often incomplete, thus requiring semantically rich models in order to be properly represented.

Even if a number of projects have addressed these topics and several applications have been tested, much work still have to be done, not only form a technical perspective, but also from a methodological one. The definition of standards both from the GeoWeb and the Semantic Web is essential because it provides the basic bricks with which building the new architectures; and the field of Cultural Heritage is advantaged because it can make use of a specific ISO standard, the CIDOC CRM. However, a lot of issues still remain, such as the difficulty of migrating to new mind sets/modelling techniques and the refinement of specific modeling aspects.

The “Ancient Milan” project aims to represent a testbed for these new methods and technologies because it offers a well defined yet very rich context. Preliminar work and the definition of the main requirements has already been already done, but it is likely that the scenario will constantly and rapidly evolve on the basis of both the results that progressively will be acquired and the new perspectives that today are not predictable.

For this reason the project has to be considered on a long term perspective and the progressive validation of the results will help in defining the validity of the proposed approach.

Acknowledgments

The author would like to thank the partners of the project for their collaboration, and, in particular Prof. M. Martini, director of CUDaM, and his staff; Prof. S. Bandini, director of CSAI, and her staff; all the researchers and collaborators at the Archaeological Museum of Milan.

References

[1] L. Aroyo, E. Hyvönen, and J. van Ossenbruggen, editors. *Proceedings of the First International Workshop on Cultural Heritage on the Semantic Web - 6th International Semantic Web Conference*, 2007.

[2] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, 2001.

[3] C. Binding, K. May, and D. Tudhorpe. Semantic interoperability in archaeological datasets. data mapping and extraction via the CIDOC CRM. In B. Christensen-Dalsgaard, D. Castelli, B. Ammitzboll Jurik, and J. Lippincott, editors, *Proceedings of the 12th European Conference on Research and Advanced Technology for Digital Libraries, Aarhus, Denmark, September 14-19, 2008*, volume 5173 of *LNCS*, pages 280–290, Berlin Heidelberg, 2008. Springer-Verlag.

[4] A. Bonomi, G. Mantegari, A. Mosca, M. Palmonari, and G. Vizzari. A system supporting users of cultural resource management semantic portals. In R. Basili and M. T. Pazienza, editors, *Proceedings of AI*IA 2007: Artificial Intelligence and Human-Oriented Computing, 10th Congress of the Italian Association for Artificial Intelligence*, volume 4733 of *LNCS*, pages 757–764, Berlin Heidelberg, 2007. Springer-Verlag.

[5] A. Bonomi, G. Mantegari, and G. Vizzari. A framework for ontological description of archaeological scientific publications. In G. Tummarello, P. Bouquet, and O. Signore, editors, *SWAP 2006 - Semantic Web Applications and Perspectives, Proceedings of the 3rd Italian Semantic Web Workshop*, volume 201 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2006.

[6] J. D. Bossler, editor. *Manual of Geospatial Science and Technology*. Taylor & Francis, London, 2002.

[7] D. Caporusso, M. T. Donati, S. Masseroli, and T. Tibiletti. *Immagini di Mediolanum*. Civiche Raccolte Archeologiche e Numismatiche di Milano, Milano, 2007.

[8] G. Christakos, P. Bogaert, and M. Serre. *Temporal GIS: Advanced Functions for Field-Based Applications*. Springer, New York, 2002.

[9] D. L. Clarke, editor. *Spatial Archaeology*. Academic Press, London, 1977.

[10] J. Conolly and M. Lake. *Geographical Information Systems in Archaeology*. Cambridge University Press, Cambridge, 2006.

[11] N. Crofts, M. Doerr, T. Gill, S. Stead, and M. Stiff. Definition of the CIDOC Conceptual Reference Model. Technical report, ICOM/CIDOC Documentation Standards Group - CIDOC CRM Special Interset Group, May 2008.

[12] F. M. De Paoli and G. Mantegari. Semantic profiling to support multi-view and multimodal interaction. In *Proceedings of the ARCHAIA Training Seminars on research Planning, Characterisation, Conservation and Management in Archaeological Sites*, British Archaeological Reports, Oxford, in press.

[13] M. Doerr. The CIDOC CRM. An ontological approach to semantic interoperability of metadata. *AI Magazine*, 24(3):75–92, 2003.

[14] S. Dragićević. The potential of Web-based GIS. *Journal of Geographical Systems*, 6:79–81, 2004.

[15] S. Erle, R. Gibson, and J. Walsh. *Mapping Hacks: Tips and Tools for Electronic Cartography*. O'Reilly Media, Sebastopol, 2005.

[16] M. F. Goodchild. Citizens as sensors: Web 2.0 and the volunteering of geographic information. *GeoFocus*, 7:8–10, 2007.

[17] M. Hildebrand, J. van Ossenbruggen, and L. Hardman. /facet: A browser for heterogeneous semantic web repositories. In I. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, M. Uschold, and L. Aroyo, editors, *Proceedings of the 5th International Semantic Web Conference*, volume 4273 of *LNCS*, pages 272–285, Berlin Heidelberg, 2006. Springer–Verlag.

[18] I. Hodder and C. Orton. *Spatial Analysis in Archaeology*. Cambridge University Press, Cambridge, 1976.

[19] D. F. Huyn, D. R. Karger, and R. C. Miller. Exhibit: Lightweight structured data publishing. In *Proceedings of the 2007 International World Wide Web Conference*, 2007.

[20] L. Isaksen. The translation framework for archaeological excavation data: Transparent negotiation and sharing of local application terminologies, instances and ontologies. Progress report submitted for continuation towards a PhD - University of Southampton, Faculty of Engineering and Applied Science, School of Electronics and Computer Science, July2008.

[21] K. L. Kelly, T. A. Pardo, S. S. Dawes, A. DiCaterino, and W. Foderingham. Sharing the costs, sharing the benefits: the NYS GIS cooperative project. Project Report 95–4, New York State Department of Environmental Conservation - Center For Technology in Government, 1995.

[22] L. D. Koontz. Geographic information systems. challenges to effective data sharing. Testimony Before the Subcommittee on Technology, Information Policy, Intergovernmental Relations and the Census, Committee on Government Reform, House of Representatives, United States General Accounting Officr, June 2003.

[23] S. Madry. An evaluation of Google Earth for archaeological exploration and survey. In J. T. Clark and E. M. Hagemeister, editors, *Digital Discover. Exploring New Frontiers in Human Heritage. Proceedings of the XXIV CAA Conference. Fargo, USA, April 2006*, pages 303–311, Budapest, 2007. Archaeolingua.

[24] D. Moretz. Internet GIS. In S. Shekar and H. Xiong, editors, *Encyclopedia of GIS*, Springer Reference, pages 591–596. Springer, New York, 2008.

[25] Z. R. Peng and M. H. Tsou. *Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Network*. Wiley, Hoboken, 2003.

[26] S. Putz. Interactive information services using World-Wide Web hypertext. Xerox PARC electronic publication, 1996.

[27] A. Scharl and K. Tochtermann, editors. *The Geospatial Web. How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society*. Springer-Verlag, London, 2007.

[28] G. Schreiber, A. Amin, M. van Assem, V. de Boer, L. Hardman, M. Hildebrand, L. Hollink, Z. Huang, J. van Kersen, M. de Niet, B. Omelayenko, J. van Ossenbruggen, R. Siebes, J. Taekema, J. Wielemaker, and B. Wielinga. MultimediaN E-Culture demonstrator. In I. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, M. Uschold, and L. Aroyo, editors, *Proceedings of the 5th International Semantic Web Conference*, volume 4273 of *LNCS*, pages 951–958, Berlin Heidelberg, 2006. Springer–Verlag.

[29] N. Shadbolt, T. Berners-Lee, and W. Hall. The Semantic Web revisited. *IEEE Intelligent Systems*, 21(3):96–101, 2006.

[30] A. Turner. *Inroduction to Neogeograpby*. O'Reilly Media, Sebastopol, 2006.

[31] D. Wheatley and M. Gillings. *Spatial Technology and Archaeology. The Archaeological Applications of GIS*. Taylor and Francis, London, 2002.

3D Digitization: making it easier and extending it to color

Roberto Scopigno*

Istituto di Scienza e Tecnologie dell'Informazione (ISTI-CNR), Pisa, ITALY

ABSTRACT

The easy construction of detailed and accurate 3D models is becoming a reality by the increasing diffusion of 3D scanning technology. The reduction in cost of the scanning devices and the increasing availability of good processing tools (including emerging open source solutions) makes 3D scanning an enabling technology for the construction of shape models. The talk will present the capabilities of this technology, presenting some recent advances (low cost scanning systems, 3D-from-images technology, improved automation of sampled data processing) and highlighting some open problems. A major focus will be how color or surface reflection characteristics could be sampled and associated with reconstructed 3D shape models. The different approaches proposed will be reviewed, giving more emphasis to the more practical solutions for both acquiring color or surface reflection and mapping those data efficiently on surface meshes. Some examples of the results of current projects, mainly in the Cultural Heritage field, will be shown.

Keywords: 3D scanning, sampled data processing, color acquisition and mapping, cultural heritage applications

Index Terms: I.3.7 [Three-Dimensional Graphics and Realism]: Color, shading, shadowing, and texture;

1 INTRODUCTION

3D technology is nowadays in a consolidate status, since 3D data can be managed on any low-cost computer, thanks to the impressing improvement of technology brought us by the huge 3D computer games market. Any PC comes equipped with everything is needed to manage interactive 3D graphics. New technologies also exist for sampling 3D shapes, usually called 3D scanners. The last ten years have shown an impressive progress of 3D scanning solutions, including both hardware devices (used to sample real objects and to return us sampled 3D point clouds) and graphics software needed to transform those sampled point clouds into good-quality 3D models and to use them in real applications.

Nevertheless, we still miss a significant impact of 3D graphics on Cultural Heritage (CH) applications. Even if we have a series of good practices and some important examples where digital 3D data played an important role, the adoption of those technologies is still far below what we could expect. There are some reasons for that: the 3D graphics field only recently reached a consolidated status; most of the experiences done so far were often driven by academia, rather than being driven directly by CH operators. Some miss concepts or wrong beliefs are also responsible of a very slow diffusion and some skepticism among our CH colleagues. Finally, color acquisition and management on scanned 3D models has been perceived as largely unsatisfactory by art experts, used to the high quality photographic medium. We will try to discuss in this paper some of the more common beliefs, with the aim of demonstrating that some of the perceived cons of this technology are due to problems which have been solved recently.

*E-mail: r.scopigno@isti.cnr.it

2 IS 3D SCANNING A TOO EXPENSIVE TECHNOLOGY?

One of the criticism more often raised against the adoption and massive use of 3D scanning technologies is the cost for the deployment of that type of technology. Especially when considering the low budget which characterizes most CH-related activities, 3D scanning cost is often perceived as excessive. Cost issues are raised at several different levels: cost of the hardware required, i.e. of the specific 3D scanning devices; cost of the software needed to process the raw data produced by 3D scanners and processing time (including also the personnel cost, since in some cases a highly skilled operator is still required); and levels of skills required for the operator to ensure proper and successful use of these technologies.

2.1 Cost of hardware

A large number of different 3D digitization devices have been proposed in the last 20 years [1]. A common distinction is between active optical vs. passive optical devices. Active optical systems can be further divided into: triangulation-based systems (using either laser or structured light patterns), and time-of-flight (TOF, also called LIDAR). Passive optical include: silhouette-based systems, stereo and multi-stereo matching solutions (which reconstruct 3D geometry from streams of photographs or videos). Currently, the most diffuse systems are active optical systems (triangulation systems for small/medium scale artifacts and TOF device for large scale artifacts, such as architecture). Unfortunately, the reduction in the cost of these systems was nearly negligible in the last ten years, much slower than the cost reduction experimented in other information technology fields. The price tag of good devices is still in the order of 30-60 KE for triangulation-based systems and 70-100 KE for TOF systems. The slow technical advance and the minor price cut is due to the fact that 3D scanning is still a niche market: since the most successful devices sell a few hundred units per year, there are not sufficient revenues for massive R&D effort and large scale production savings. For small/medium scale acquisition, the recent introduction of a low-cost laser-based device sold at 2500 USD is a remarkable news, which should have a giant impact on the domain (it is a triangulation-based system, see at https://www.nextengine.com/). A similar impressive reduction of TOF cost is still a dream, but the good news is that the acquisition of large scale artifacts can now be approached by adopting new passive optical methodologies, in particular the ones that perform 3D reconstruction from a simple sequence of high resolution digital photos of the artifact [9, 12]. These methods are an evolution of the old photogrammetry approach, they have been considerably improved recently and show some interesting potential for a very wide diffusion. They are based on the search of a small set of correspondences between the processed images; these correspondences (usually in the order of tens or one hundred) identify some feature points in the scene as seen from different point of view. Depending on how these corresponding image points are located in the different pictures, the 3D position of these feature points and the orientation of the camera are recovered. Starting from these few sparse points, a dense depth range map can be reconstructed from each image by interpolating these recovered points and applying stereo-matching techniques on the pixel in the in-between regions. An example of result obtained with this technology is shown in Figure 1, where the model presented has been reconstructed by processing

Figure 1: Image from a 3D model obtained with passive reconstruction from a set of digital images (by ARC 3D and MeshLab tools).

some photos, shot all around the statue, using the ARC 3D Webservice (http://www.arc3d.be/) developed within the EC IST Network of Excellence "Epoc" (http://www.epoch-net.org). The raw data returned by the ARC 3D system have been processed with the MeshLab tool (http://meshlab.sourceforge.net) [5] .

The advantages of this new approach are quite evident. The only hardware required is a simple good quality digital photographic camera, and the scanning process requires just taking a reasonably large number of photos all around the object. On the other hand, this approach still exhibits a geometric precision that is much less predictable than the well assessed laser-based 3D scanning technologies: since the reconstruction process is based on the detection of corresponding features on consecutive photos, these approaches encounter difficulties in the reconstruction of artifacts with large flat and uniformly colored parts that do not exhibit evident features to be recovered (e.g. uniformly painted walls) and have even more significant problems with non-diffuse surfaces.

2.2 Cost of software

Unfortunately, 3D scanning systems do not produce a final, complete 3D model but rather a large collection of raw data, which have to be post-processed. A complete scan of an artifact requires the acquisition of many shots taken from different viewpoints to gather complete information on its shape. Each shot produces a range map, that is a single partial view of the object. The number of range maps required to sample an artifact depends on the surface extent of the object and on its shape complexity. Usually we sample from a few tens up to a few hundred range maps. Range maps have to be processed to convert the data encoded into a single, complete, non redundant, and optimal digital 3D representation (usually, encoded by a triangulated surface). The processing phases (usually supported by commercial tools) are:

- Range Maps Alignment. By definition, the range map geometry is relative to the current sensor location and has to be transformed into a common coordinate space where all the range maps lie well aligned on their mutual overlapping regions (i.e. the sections of two adjacent range maps which sample the same portion of the artifact surface).

- Range Maps Merge (or Reconstruction). A single, non-redundant triangulated mesh is built out of the many partially overlapping range maps. This processing phase reduces the redundancy (after merging, each surface parcel of the artifact will be represented by just one geometric element).

- Mesh Editing. The goal of this step is to improve (if possible) the quality of the reconstructed mesh, for examples, reducing noisy data or fixing un-sampled regions (generating surface patches to close small holes).

- Mesh Simplification. The huge complexity of the model obtained usually has to be reduced in a controlled manner to transform the usually huge master model (millions of samples and triangles) into a model of size appropriate to the specific application. A huge mesh can be either simplified or converted into a discrete a Level-Of-Detail (LOD) model or a multiresolution representation.

- Color Mapping. The information content is enriched by adding color information (an important component of the visual appearance) to the geometry representation.

All these phases are supported either by commercial (e.g. INUS Technology, InnovMetrics, Raindrop Geomagic) or academic tools [10, 3, 5] . Unfortunately, commercial software is still very costly (around 10K-20K Euro for each installed workstation). The diffusion of open source solutions could be an important resource for fostering an increased diffusion of this technology; some academic labs are following this policy [5].

2.3 Time required to process the raw data

The results of the last ten years of research on sampled 3D data processing had a profound impact on the time and effort requested to the user to transform the raw, point-based sampled data into a good quality 3D model. Processing large sampling was a nightmare until recently. Taking into example the case of a single statue, processing time has been reduced from several weeks to a few days (1-3), thanks to a progressive automation of the process. Improved management of a really large set of range maps (from 100 up to 1000) can be obtained both by providing a hierarchical organization of the data (range maps divided into groups with atomic alignment operations applied to an entire group rather than to the single scan) and by using a multiresolution representation of the raw data, to make rendering and processing more efficient. Moreover, since the standard approach (user-assisted selection of each overlapping pair and selection of the correspondent alignment pairs) becomes impractical on a large set of range maps, some solutions for a completely automatic range map alignment have been proposed. These methods are based on the characterization of a few feature points contained in each range map and subsequent search for matching points in the adjacent maps. These solutions have been demonstrated to work well (90-95% reduction of the processing time) but are still available only in academic solutions [7, 3]. Scanning systems that automatically track the scanner location and therefore produce aligned range maps also exist (based on magnetic or optical tracking), but usually cost twice than standard high quality devices and are thus of limited diffusion.

3 Is 3D scanning limited to geometry sampling?

Most 3D scanning systems consider just the geometric shape acquisition, while a very important aspect in CH applications is color sampling. This is the weakest feature of contemporary technology since those scanners that acquire color information usually produce low-quality color sampling (with a notable exception of the technology based on multiple laser wavelengths, unfortunately characterized by a very high price). Moreover it should be noted that existing devices sample only the apparent color of the surface (reflected color) and not its reflectance properties, which constitute the characterizing aspect of the surface appearance. The availability of a digital model encoding how a given surface reflects the light is of extreme importance if we would be able to see the digital 3D replica under different lighting conditions. Let me introduce here just a few

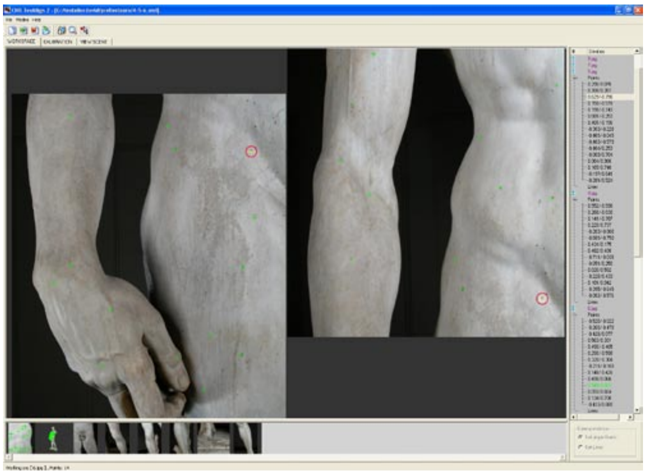


Figure 2: Screenshot of our Image Alignment tool with an example of partially overlapping RGB images (red circles indicate an image-to-image correspondence).

Figure 3: Two set of around 60 images each (depicting the pre- and post-restoration status) have been mapped onto the digital model of Michelangelos David and rendered in real time using the Virtual Inspecor system. Digital model courtesy of Stanford University (Digital Michelangelo Project) and Museo Gallerie dellAccademia, Florence.

examples to justify the need of dynamic illumination capabilities: being able to move a light source interactively around the digital replica and to synthesize accurately how the object is lighted, e.g. simulating razing light; reproducing different daylight conditions on an architecture; or simulating the visual appearance of an artifact or an architecture when it is lighted with different illuminants (electric light, different type of flames, solar light, etc).

Several accurate approaches for sampling the surface reflection characteristics have been proposed; a majorexample is the methodology devised by MPII to acquire the Bidirectional Reflection Function Distribution (BRFD) [11]. Unfortunately, most of these solutions are still too complicated to be massively applied to the CH field, where it is very hard to setup the controlled lab conditions needed to estimate the light reflection and diffusion. Moreover, since these methods makes use of controlled lighting conditions (usually in lab conditions) to sample the reflection function, they are nearly impossible to use on architectures.

For most practical cases a simpler approach is still widely used: the so-called apparent color is acquired and mapped to the 3D model. A series of pictures can be taken with a digital camera, trying to avoid shadows and highlights by taking them under a favorable lighting setup; these photographs are then stitched onto the surface of the object. However, even in this simpler case, the processing needed in order to build a plausible texture is not straightforward [4]. Naive mapping of apparent color on the mesh can produce severe discontinuities that are due to the varying illumination over the surface sampled by the photos. Some approaches have been proposed to reduce the aliasing and to produce seamless color mapping. A new flexible solution has been proposed in [2], where a multivariate blending function weights all the available pixel data with respect to geometric, topological and colorimetric criteria. The blending approach is efficient, since it mostly works independently on each image, and can be easily extended to include other image quality estimators. The resulting weighted pixels are then selectively mapped on the geometry, preferably by adopting a multiresolution per-vertex encoding to make profitable use of all the data available and to avoid the texture size bottleneck.

A basic problem in managing color information is how to register the images with the geometric data. In most cases, the set of images is taken after the scanning, using a consumer digital camera. This registration step is again a complicated time-consuming phase which requires substantial intervention of a human operator. Unfortunately, no fully automatic and robust approach has been proposed for the general problem (i.e. a large and complex object, where each image covers only a subset of its overall extent). The user is usually required to provide correspondences, or hints on the correspondences,which link the 2D images and 3D geometry (see Figure 2).

In a recent research we designed a new tool to support *image-to-geometry alignment*, *TexAlign* [8], whose main goals were: to reduce the user intervention in the process of registering a set of images with a 3D model; to improve the robustness of the process by giving the user the possibility of selecting correspondences which link either 2D points to 3D geometry (*image-to-geometry* correspondences) or 2D points to 2D points (*image-to-image* correspondences). The latter can help a lot in all those cases were a single image covers a region where the surface has not sufficient shape feature to allow an accurate selection of *image-to-geometry* correspondences. The *TexAlign* tool tries to solve the problem by setting up a *graph of correspondences*, where the 3D model and all the images are represented as nodes and a link is created for any correspondence defined between two nodes. This graph of correspondences is used to keep track of the work done by the user, to infer automatically new correspondences from the one instantiated and to find the shortest path, in terms of the number of correspondences that must be provided by the user, to complete the registration of all the images.

In all those cases where the operator has a large number of images to align and map to the 3D shape, *TexAlign* allows to reduce the time needed to perform the alignment and to improve the overall accuracy of the process. Some results are reported in [8]. This system has been recently used to map a complex photographic sampling (more than 70 images to be mapped on the David model, see Figure 3) and [6].

Considering the various technologies and methodologies used for 3D digitization, the subset of techniques for surface reflection acquisition and mapping on digital 3D models is the topic where greater is the potential for improvement to cope with the pressing requirements of CH applications.

Some results of high-quality mapping of color data on 3D meshs are presented in Figures 3 and 4.

Figure 4: An example of colored 3D digital model of one of the terrecotte that decorated the front of the Luni temple. On the left, the current color and on the right, a preliminary hypothesis of the original painted status.

4 CONCLUSIONS

As briefly presented in the previous sections, we think that the evolution and improvement of 3D scanning technologies makes this approach highly effective for applications in the CH domain. This technology is now affordable and satisfies the data accuracy and density required by many applications. We forecast a wide adoption in the near future. What still remains, in parallel with the further improvement of the technology (3D sampling HW, SW for geometric post-processing), is the required management of metadata and provenance data, which should be archived and managed through the entire workflow of geometric data.

ACKNOWLEDGEMENTS

We acknowledge the financial support of project *"START: Scienze e Tecnologie per il Patrimonio Artistico, Architettonico e Archeologico Toscano"* financed by Regione Toscana, Italy.

REFERENCES

[1] F. Blais. A review of 20 years of range sensor development. In *Videometrics VII, Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 5013*, pages 62–76, 2003.

[2] M. Callieri, P. Cignoni, M. Corsini, and R. Scopigno. Masked photo blending: mapping dense photographic dataset on high-resolution sampled 3d models. *Computers and Graphics*, 32:464–473, 2008.

[3] M. Callieri, P. Cignoni, F. Ganovelli, C. Montani, P. Pingi, and R. Scopigno. VCLab's tools for 3D range data processing. In A. C. D. Arnold and F. Niccolucci, editors, *VAST 2003*, pages 13–22, Bighton, UK, Nov. 5-7 2003. Eurographics.

[4] M. Callieri, P. Cignoni, and R. Scopigno. Reconstructing textured meshes from multiple range rgb maps. In *7th Int.l Fall Workshop on Vision, Modeling, and Visualization 2002*, pages 419–426, Erlangen (D), Nov. 20 - 22 2002. IOS Press.

[5] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia. Meshlab: an open-source mesh processing tool. In *Sixth Eurographics Italian Chapter Conference*, pages 129–136. Eurographics, 2008.

[6] M. Dellepiane, M. Callieri, F. Ponchio, and R. Scopigno. Mapping highly detailed color information on extremely dense 3D models: the case of Davids restoration. *Computer Graphics Forum*, 27(8):(in press), 2008.

[7] A. Fasano, P. Pingi, P. Cignoni, C. Montani, and R. Scopigno. Automatic registration of range maps. *Computer Graphics Forum (Proc. of Eurographics '05)*, 24(3):(in press), 2005.

[8] T. Franken, M. Dellepiane, F. Ganovelli, P. Cignoni, C. Montani, and R. Scopigno. Minimizing user intervention in registering 2D images to 3D models. *The Visual Computer*, 21(8-10):619–628, sep 2005.

[9] R. Koch, M. Pollefeys, and L. V. Gool. Realistic surface reconstruction of 3D scenes from uncalibrated image sequences. In F.-E. Wolter and N. M. Patrikalakis, editors, *The Journal of Visualization and Computer Animation*, volume 11(3), pages 115–127. John Wiley & Sons, Ltd., 2000.

[10] S. C. G. Laboratory. Vrippack volumetric range image processing package. More info on: `http://www-graphics.stanford.edu/software/`, 2003.

[11] H. Lensch, J. Kautz, M. Goesele, W. Heidrich, and H.-P. Seidel. Image-based reconstruction of spatial appearance and geometric detail. *ACM Transaction on Graphics*, 22(2):234–257, Apr. 2003.

[12] M. Pollefeys, R. Koch, M. Vergauwen, A. A. Deknuydt, and L. J. V. Gool. Three-dimensional scene reconstruction from images. In B. D. Corner and H. Nurre, Joseph, editors, *Proceedings of the Conference on Three-Dimensional Image Capture and Applications II (3d-image-capture-appl-00)*, pages 215–226, Bellingham, Washington, Jan. 24–25 2000. SPIE.

Intangible heritage management and multimodal navigation

R. Schettini, G. Ciocca, A. Colombo, I. Gagliardi,*

D.I.S.Co. (Dip. Informatica, Sistemistica e Comunicazione)
Università degli Studi di Milano-Bicocca,
Viale Sarca 336, 20126 Milano Italy
Schettini, colomboal, ciocca @disco.unimib.it

** ITC, Consiglio Nazionale delle Ricerche*
Via Bassini 15, 20133 Milano, i.gagliardi@itc.cnr.it

Abstract— **In this paper we present the tools and the methods which have been developed in order to manage and consult multimedia ethnographical archives, the content of which is composed of text, images, audios (both songs and spoken documents) and videos. The system offers the user several retrieval strategies for querying the multimedia archive database exploiting alphanumeric relational query, audio similarity query and clustering, image and video similarity. Once a subset of materials meeting the user's information needs have been identified, these images can be displayed in a 3D virtual exhibition which the user can visit interactively. The system presented is actually exploited to manage and multimodal navigate the Archive of Ethnography and Social History of the Lombardy Region some 18,000 oral documents, 3,000 textual transcriptions, 2,000 musical transcriptions, 18,000 MP3 audio files, 10,000 photos, 500 videos [1].**

I. Introduction

According to the 2003 Convention for the Safeguarding of the Intangible Cultural Heritage of Unesco, the intangible cultural heritage (ICH) – or living heritage – is defined as the practices, representations, expressions, as well as the knowledge and skills, that communities, groups and, in some cases, individuals recognise as part of their cultural heritage. The Convention states that the ICH is manifested, among others, in the following domains:

1. Oral traditions and expressions including language as a vehicle of the intangible cultural heritage;
2. Performing arts (such as traditional music, dance and theatre);
3. Social practices, rituals and festive events;
4. Knowledge and practices concerning nature and the universe;
5. Traditional craftsmanship.

The Archive of Ethnography and Social History of the Lombardy Region (AESS) was founded to preserve, study, and enhance the value of documents and images of the life, social transformations, literature, oral history, material culture, and anthropic landscapes of the Lombard territory.

The archive is composed of 18.000 oral documents, related to songs, 3000 textual transcriptions, 2000 musical transcriptions, 5000 audio files in MP3 format, and 10000 photographic documents that is related to photographs. It is managed through a database which integrates the catalogue cards with multimedia objects of different types: audio files, images, digital videos, textual transcriptions, musical scores, etc…

In this paper we present the tools and the methods which have been developed in order to manage and consult multimedia ethnographical archives, the content of which is composed of text, images, audios (both songs and spoken documents) and videos.

II. The AESS Web Site

The first requisite considered in designing the interface of the AESS web-site, was usability, defined, according to the ISO standard 9241, as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". To implement a functional, efficient and effective web site, we have used a quality model of simple employ in the design phase, so that we could evaluate the site quality in a structured way, applying six fundamental criteria, and taking into account site scope, users and context of use.

Content: does the site information content meet the objective? Is the information relevant? Complete? Reliable? Updated? Functionality: are the site functionalities adequate to its objective? Do they function correctly? Management: is the site correctly managed? Communication: does the homepage immediately render the aim of the site? Does the site correctly communicate the brand of the organization? Is the style of communication consistent with the aim of the site? Usability: is the site usable? Accessibility: is the site easily and quickly accessed? Is it easily reached by the most common search engine? Is it adequately referenced by other sites? Is the URL easily remembered? Is the site independent from the browser? Can it be accessed by disabled users?

These criteria were applied both when designing the interface and when assessing the implemented web site. The main W3C guidelines have been followed, to ensure good accessibility [2].

A. Types of Information

The archive of the AESS web site stores information concerning the oral history of the Lombard region: the data concerns mainly popular songs and other audio records describing the popular traditions handed down generation by generation, such as traditional fairs, and customs. The images and videos represent occasions in which the audios have been performed as songs, recorded as interviews, etc. Linked with the audio and image data are books, journal, discs, DVD, etc, on which these information are stored (printed, recorded, etc.)

Four different types of information, which can be indexed, searched and retrieved in each query session, have been identified:

1. oral documents: these form the core of the archive, and consist of cards describing each item by title, incipit, metric, keywords, coupled with audio files, and also textual and musical (pentagram) transcription.
2. devices: physical devices (books, discs, CDs, etc.) on which the document is available
3. events: occasions in which audio documents have been recorded, or photographs taken.
4. images: photographs or videos representing events during which audios have been recorded.

Every item has a textual card describing it, with information such as the title, author, date, some keywords taken from a manually composed dictionary, the same for all the items, and, if possible, a description of the object (e.g. in the photographs).

B. Users

Three main kinds of users have been identified:

Lombardy Region staff: directors and technicians, who have competences and requirements peculiar of their own work; professional users: experts, ethnographer, and art historians who need a simplified interface in order to query the system, but are accustomed to the specific terms used in the cards; generic users without any competence neither in ethnography nor in browsing the web and therefore require a further simplified query interface.

C. Multimodal Navigation and Retrieval

The design of the database and web site provides facilities that allow all users, even if not expert in the field, or unfamiliar with the database contents, or with the language in which database terms are expressed, to query the database successfully.

Multimodal means of navigation and retrieval of the different kinds of information have been designed and implemented:

- retrieval in SQL standard, on different cards, designed to allow the different users to interact with the system: four different kinds of search are offered.
- similarity-based textual description: once the user has obtained a list of documents, in response to a query, he can apply the “similarity link” algorithm to retrieve the most similar objects according to their description;
- similarity image retrieval: the user can perform a similarity search in order to retrieve images similar in pictorial contents.
- similar audio retrieval: the user can navigate among the most similar audio documents according to their acoustic similarity;
- clusters of audio files: the user can choose among the available audio files, clustered by the algorithm defined above. The clusters obtained for the AESS audio archive are 83 with less than ten items in each.

The user can query the archive through different interfaces, described here in crescent order of interaction with the system:

- Catalogue: it displays the whole content of the archive, allowing the navigation among the four types of information (oral documents, devices, events, and images).
- Guided search: it provides the users with predefined query as “Carnevale di Bagolino” or “Musica delle Quattro province” with a brief explanation.
- Simple search: it supplies a simplified query interface in which the user is required to specify what, who, how and where to retrieve information from the archive.
- Advanced search: made up of four query forms, each one for a specified information type, it allows querying separately each type of information. By filling a field of a form (e.g. “device”), it will be returned all the devices which satisfy the query.
- Search by examples: the result of the image similarity search can be iteratively refined by applying a relevance feedback mechanism based on image examples submitted by the user.
- Specialized search: a contained of innovative search tools. At the moment it contains Cluster audio functionality.

Figs. 1-6 show some screen shots of the multimodal navigation.

The AESS website implements, besides the standard functionalities of catalogue and search, various modalities of navigation and employ. These include, among others, the guided tours, and the retrieval and clustering of similar audios that is the organization of the audio files stored in the database in groups, containing files acoustically similar to each other. These functionalities will be described in details in the following sections.

Fig. 1: The home page of AESS web site


Fig. 4: Simple search form

Fig. 2 "Carnevale di Bagolino" guided search

Fig. 5: Audio document card

Fig. 3: Images retrieved by "Carnevale di Bagolino" guided search

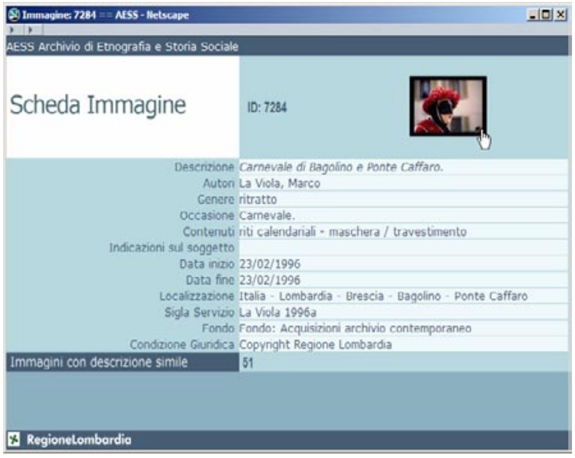


Fig.6: Image card

At present, two guided tours - "Canto narrativo" (narrative song) and "Piffero" (pipe or fife)- have been finalized, while two ones more, about rituality and folk performance, are been added soon. These guided tours are described through a selections of songs, performance, interviews, video clips, etc., recorded or gathered in North Italy, that the user can listen to, read, and compare to completely explore the content.

Fig. 7 shows the AESS web site structure.

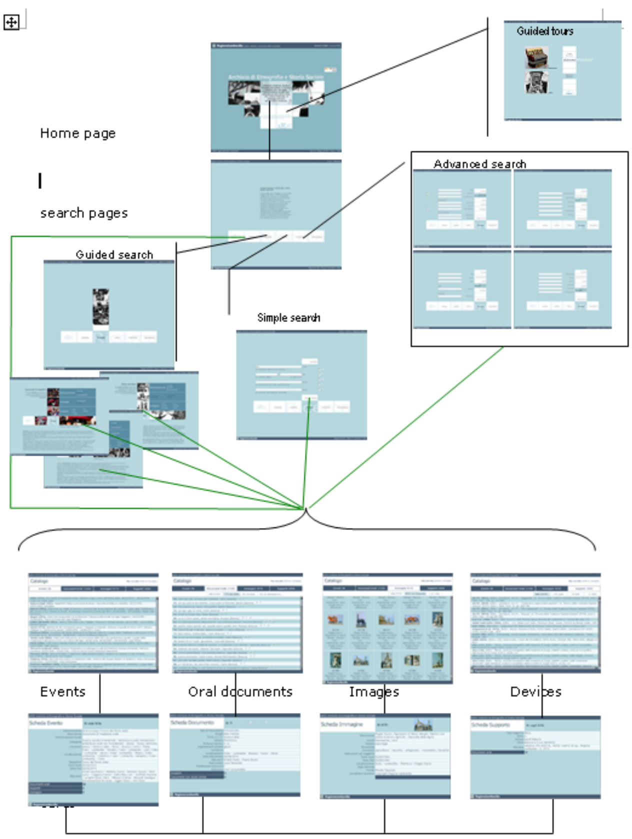


Fig. 7: AESS web site structure

III. Tools and Methods in the AESS site

A. Similarity-based textual description

Audio, video and images are often accompanied by a textual description of their semantic contents. The basic idea is that the presence of terms common to two different texts indicates that the objects described can be considered similar to each other. The texts can be used in indexing and retrieving objects if significant terms are identified by a dictionary, and a suitable similarity function among sets of significant terms is defined [[3], [4]].

The dictionaries employed contain sets of significant terms that can be used to index textual annotations; they can be created either automatically, as the result of an IR process using lists of stop-words (terms, such as articles or adverbs, that are "poor discriminators", too frequent to be significant), or manually, by experts in the domain, who indicate the more significant terms according to the criteria applied.

Our system, designed for general use, can utilize either automatically or manually created dictionaries. In the former case they include all the terms present in the textual annotations (excepting those on a standard Italian stop list). No stemming procedure is applied, as no satisfactory algorithm is available for the Italian language.

Most morphological variations (singular/plural, feminine/masculine, ...) are, however, automatically eliminated.

We advise the use of a controlled dictionary of the terms present in the annotations if available. A weight can be assigned automatically to each term, on the basis of the number of times the term occurs in the entire collection, following well-established procedures of Information Retrieval (IR), or manually, considering the importance of the term in the domain, or in the collection, regardless of its frequency [5].

The algorithm that calculates the degree of similarity between the text of the query object and each of the other texts contained in the database is an extension of Salton's [6] well known formula which takes into account the weight wi of the i-th descriptor concerned [7]:

$$sim_{i,j} = \frac{2(w_i term_i \cap w_j term_j)}{w_i term_i \cup w_j term_j}$$

and where $sim_{i,j}$ can assume any value in the range of [0, 1]. The greater the value of $sim_{i,j}$, the greater the similarity between the two textual annotations.

B. Similar audio retrieval tools

When a user, applying any type of consultation selected, reaches the information card of any oral document, links to possible similar documents from the acoustic point of view are also shown (Fig. 8).

The acoustic similarity among various oral documents (especially if musical) of the database has been computed with the TreeQ system [8], implemented by Jonathan T. Foote of the Institute of Systems Science, at the National University of Singapore [9]. This method represents each audio file as a histogram which encodes some fundamental physical features of the file.

Fig. 8. Card of an audio document with the "acoustically similar" documents shown.

These histograms can be considered vectors; therefore, the acoustic similarity index between two files is estimated computing the distance cosine between the related vectors: the closer the index is to 1, the more the two file are similar, in their acoustic features (Fig. 9).

Once the similarity between all the possible couples of documents has been computed and the meaningful results (that is, all the couples with a similarity index greater than 0.8) have been listed in a table of the database, the results are presented to the user.

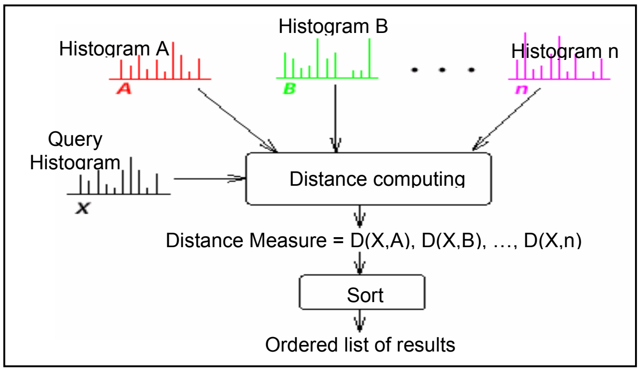


Fig. 9: Similarity computing among audio files

C. *AUDIO CLUSTERING Tool*

The AESS website also implements the functionality of audio clustering, supplying information about the file in a general manner, that is, not referred to a particular file.

A cluster can be simply defined as a group of similar objects and the implemented algorithm summarized in few words: given n vectors, the algorithm divides them into K groups, or clusters, so that the vectors belonging to the same cluster are more similar to each other than to vectors belonging to the other clusters [10].

As in the traditional clustering processes, the division of the vectors is reiterated until certain conditions are satisfied; in our case the process is interrupted when, inside each cluster, the similarity index among all the possible couples of vectors is greater than a threshold empirically set at 0.8.

According to this criterion, we have developed an algorithm that functions in this manner [11]:

- after "randomly" selecting n vectors, a first clustering is computed which assigns all the vectors to n different groups;
- in each cluster, the algorithm searches for the couple of vectors with the lowest index of similarity: if this index is greater than the set threshold, the cluster is accepted, otherwise it is divided into two sub-clusters, choosing the two least similar vectors as centroids of the new groupings;
- a new clustering is computed on those vectors that have been assigned to the divided cluster;
- once all the clusters satisfy the evaluation criterion, the algorithm estimates whether it is possible to group together two or more different clusters to produce a new cluster that still satisfies the conditions set by the evaluation criterion.

In particular, the clustering is computed as follows:

- each file is assigned to the cluster identified by the centroid to which it has resulted most similar;
- the barycenter for each cluster is computed on the basis of the vectors belonging to it;
- the similarity between the barycenter and the centroid is then computed for each cluster. If the similarity is equal to 1 (that is, the barycenter and the centroid coincide), the process ends; otherwise the barycenter is set as a new centroid and the process is reiterated from the beginning (in other words, the clustering process terminates when two successive iterations return the same partition of the vectors into clusters).

In this manner, the number of clusters increases and decreases during the process, adjusting to the nature of the data. The final number of clusters is not influenced by the initial one, although the initial number influences the speed of the algorithm. The speed is also affected by the quality of the first clustering.

To guarantee that the initial clustering is good enough, the choice of the initial centroids is fundamental: these are selected randomly, on the sole condition that they not be too similar (their similarity index must not exceed a suitable threshold), so as to "cover" the whole vectorial space in the most uniform way. An initial clustering starting from centroids selected in this way assures a lower number of successive partitions and re-joinings.

The implemented algorithm has produced a satisfactory clustering and, above all, one that is sufficiently stable. The major problem of clustering algorithms is, in fact, that they are completely dependent on the selection of the initial centroids. Our algorithm too is affected by this selection, but the possibility of varying the number of clusters allows us to curtail that dependence drastically, rendering the solution more reliable and stable.

The result of the clustering is saved in a database table, and can be viewed from the website (Fig 10) by selecting "Specialized search" and then "Cluster Audio". Finally, when the user selects one of the cells corresponding to the clusters (colored with increasing intensity, according to the number of audios in each cluster) a new page is displayed with the data related to the audio files in the selected (Fig. 11).

Fig. 10. Page of audio files clustering

Fig. 11. Audio documents belonging to cluster no. 5.

D. Similarity-based image retrieval Tools

The similarity image retrieval module is based on the QuickLook image retrieval system [12]. The QuickLook system combines within a single framework the capabilities of alphanumeric relational query, the content-based image and video query exploiting automatically computed image features, and the textual similarity query using any textual annotations attached to database items (such as figure and video captions). The system offers the user several retrieval strategies for querying the database. He can then progressively refine the system's response by indicating the relevance, or non-relevance of the items retrieved. Its framework can be adapted to support different image categories and tasks [13]. Its functionalities are continuously extended and updated with new features.

Similarity-based image retrieval can be accomplished using pictorial attributes and (if available) textual annotations both referring to the image content. A generic query may be composed of visual and/or textual parts (sub-queries). During the retrieval phase, each sub-query is processed separately, and then the results are combined with a similarity function to obtain a final score. The images are then ranked according to this score. If the user is not satisfied by the system's response, he can refine the search modifying the query with examples of what is relevant or non-relevant to what he is looking for. The similarity function is dynamically adapted to the query by a relevance feedback mechanism [14] that modifies the similarity function used to evaluate the images. Since comparing an image query Q with every image I in the database is a time-consuming task, we have implemented a method for filtering the database before the pictorial distances are actually computed. This method is based on a variant of triangle inequality [15], and has the advantage of being applicable to any distance measure that satisfies triangle inequality.

Figs. 14-15 show an example of image similarity retrieval which exploits the relevance feedback mechanism. The user start by selecting an image card and then by clicking on the "Search similar images" link. The system provides an initial ranking of the images in the database starting from the most pictorially similar. The user can then use the interface to select the images that are really similar to the initial one (positive examples), and the images that are not similar (negative examples), and resubmit the query. The system updates the query and the similarity function and returns a refined ranking. The user can iterate through this process until the result is considered satisfactory. From this interface the user can open the image cards, and start another image similarity search.

Fig. 14. Initial image similarity result. The initial query is the top left image.

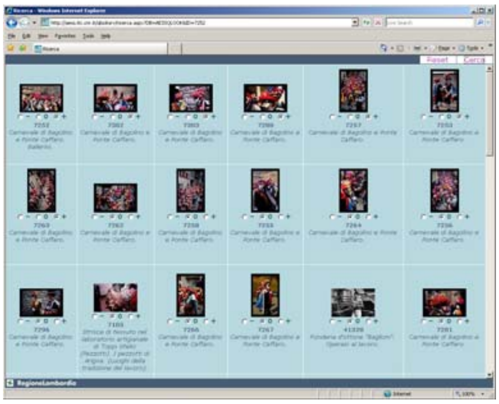


Fig. 15. Refined image similarity result.

E. Video Analysis and Summarization Tools

Video sequences are analysed and summarized (according to their visual contents) with still images. On these images, the same retrieval strategies developed for images can be applied to retrieve video sequences. The still images chosen to represent the video content are called key-frames. Automatic video analysis and indexing is a complex process that involves different tasks [16]. Video indexing must capture the spatio-temporal contents of the video in a compact way. In order to do so, the first step in video indexing is the definition of a suitable representation of the visual content (i.e. a suitable representation for the video frames). We segment the video into shots (a continuous sequence of frames taken over a short period of time) by detecting abrupt changes and fades between them, since these are more common than other editing effects. A gradual transition detection algorithm is currently being developed, and it will be integrated in a similar manner. Once the shot have been detected, key frames must be extracted from each shot. Our summarization system dynamically selects representative frames from the shots by analysing the complexity of the events presented and discarding redundant information [17]. An example of visual summary from a video sequence is shown in Fig. 17a. The set of frames in the summary can be further processed [18] in order to remove frames that are not significant, duplicates or very similar. A hierarchy of summaries, each of which consists of a different number of key frames, can be created to allow easy content navigation. An example of visual summary post-processing is shown in Fig. 17b. At each step in the video analysis process, we can store in a database different gathered information and index the video with the shots and scenes positions, their duration, the set of frames and so on. These information can then be used for retrieval purpose.

These tools have been created and made available to the AESS staff, to help them catalogue the video sequences.

a

b

Fig. 17. Example of visual summary extraction and post processing. The initial summary with many redundant and uninformative frames. b) The same summary after the processing phases). The original video contains 45,753 frames, the initial summary 55 and the final summary 19.

F. 3D Visualization Tools

We have realized a system for the generation of three-dimensional virtual environments using VRML in which digital items can be collocated, displayed, and visited. These three-dimensional environments are totally navigable and allow the end user to perceive the space, the proportions and the dimensions of the environment of exposure and of the objects within. Each environment is able to welcome different kind of multimedia objects such as images, video, audio, texts and three-dimensional objects. They are directly collocated in the environment (as in the case of images and videos) or through an appropriate simplified visual representation of the object (avatar). To this last category belong texts, audio and 3D objects. Every object has a link to a presentation card which usually consists in an HTML page that can contain informative fields such as author, year, place in which the original one is present, real dimensions, description, etc.

Fig. 19 shows an example of virtual environment. The objects are can be positioned in the environment on the floor and on the walls using different positioning approaches. The number of rooms that the virtual environment is not fixed a-priori but depends on the number of objects to be collocated.

The virtual exhibition created for the AESS web site can be seen by visiting the site:

http://aess.itc.cnr.it/museo_virtuale.htm. (Figure 20)

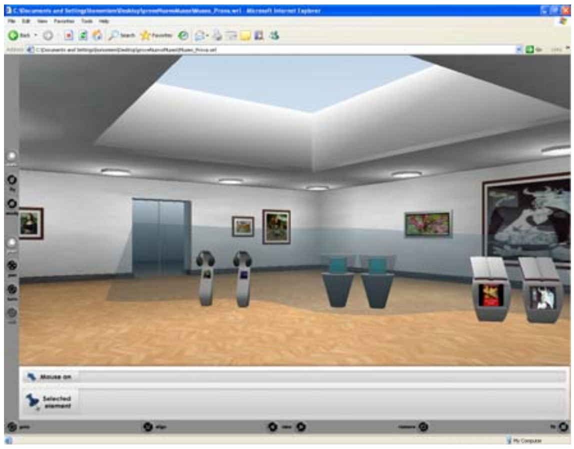


Fig. 19. Example of virtual exhibition. In the room can be seen the image positioned on the walls and avatars for audio, textual and 3D objects.

Fig. 20.The virtual exhibition created for the AESS web site.

IV. Conclusions

The AESS web site is available at the address http://www.aess.regione.lombardia.it (or http://www.aess.itc.cnr.it/)

This paper has presented some methods and tools for the retrieval and navigation of multimedia documents, (audio documents, indexed on the basis of their acoustic features, in particular) in ethnographic archives. The different approaches employed have been shown, and the advantages of their application illustrated in a real case, where multimodal navigation allows users to interact with the data in a more detailed and complete manner. Preliminary results, have proved satisfactory.

This study was performed as a part of a Regione Lombardia research contracts, and partly funded by Fondazione Cariplo projects. Thanks to the Lombard Region, and in particular to the Archivio di Etnografia e Storia Sociale for the permission to publish this paper.

REFERENCES

[1] AESS web site http://www.aess.regione.lombardia.it (2008) or http://www.aess.itc.cnr.it.
[2] W3C: Web Content Accessibility Guidelines 1.0 http://www.w3.org/TR/WAI-WEBCONTENT/
[3] Salton G., Automatic text processing, Addison-Wesley, New York, 1989.
[4] Salton G., McGill M., Introduction to Modern Information Retrieval. McGraw-Hill, 1983.
[5] Gagliardi I., Zonta B., “Comparing dictionaries in automatic construction of hypertextual links: a case study in art catalogues”, Visual Information and Information Systems (D.P. Huijsmans, Arnold W.M. Smeulders eds.), pp. 358-366, 1999.
[6] Salton G., A. Singhal, M. Mitra, C. Buckley, “Automatic text structuring and summarization”, Information Processing & Management, 33(2), pp. 193-207, 1997.
[7] Gagliardi I., Zonta B., “Quicklink: A system for the generation of similarity links in Cultural Heritage archives”, Journal of Cultural Heritage, 2(2), pp.155-162, 2001.
[8] J.T. Foote. The TreeQ Package. Available from ftp://svr- ftp.eng.cam.ac.uk/pub/comp.speech/ tools/ treeq1.3.tar.gz.
[9] Jonathan Foote “An overview of audio information retrieval”, January 1999 Multimedia Systems, Volume 7 Issue 1
[10] Jain K, Dubes R. C. Algorithms for clustering audio, Prentice-Hall, 1988.
[11] Gagliardi I., Pagliarulo P. Audio Information Retrieval in HyperMedia environment *Proceedings of the sixteenth ACM conference on Hypertext and hypermedia*, Salzburg, Austria, pp. 248 – 250, 2005
[12] Ciocca G., Gagliardi I., Schettini R, “Quicklook^2: An Integrated Multimedia system”*, Journal of Visual Languages and Computing*, **12**, pp. 81-103, 2001.
[13] Ciocca G., Schettini R., “Content-based similarity retrieval of trademarks using relevance feedback”, *Pattern Recognition*, **34**, pp. 103-199, 2001.
[14] Ciocca G., R. Schettini, “A relevance feedback mechanism for content-based image retrieval”*, Information Processing and Management*, **35,** pp. 605-632, 1999.
[15] Berman A.P., Shapiro L.G. “A flexible image database system for content-based retrieval”, Computer Vision and Image Understanding, 75, pp. 175-195, 1999.
[16] Nevenka Dimitrova, Hong-Jiang Zhang, Behzad Shahraray, M. Ibrahim Sezan, Thomas Huang, Avideh Zakhor: Applications of Video-Content Analysis and Retrieval, *IEEE MultiMedia,* **9(3),** pp. 42-55, 2002.
[17] G. Ciocca, R. Schettini, An Innovative Algorithm for Key Frame Extraction in Video Summarization, Journal of Real-Time Image Processing (Springer), Vol. 1, No. 1, pp. 69-88, 2006.
[18] Ciocca G., Schettini R., Supervised And Unsupervised Classification Post-Processing for Visual Video Summaries, IEEE Transactions on Consumer Electronics, Vol. 2, No. 52, pp. 630-638, 2006.

FIRSt: a Content-based Recommender System Integrating Tags for Cultural Heritage Personalization

Pierpaolo Basile, Marco de Gemmis, Pasquale Lops, Giovanni Semeraro, Massimo Bux, Cataldo Musto, Fedelucio Narducci
Universitá degli Studi di Bari, Dipartimento di Informatica
via E. Orabona, 4 - 70126 - Italy
{basilepp,calefato,degemmis,lops,semeraro,bux,musto,narducci}@di.uniba.it

***Abstract*—Cultural heritage personalization and Web 2.0 joint research efforts have recently emerged in the attempt to build social and collaborative approaches to solve the problem of filtering content in the context of art museums. One way to tackle the problem of recommending artifacts to visitors is to take into account not only the official textual descriptions, but also the user-generated content, namely the tags, which visitors could use to freely annotate relevant works. The main contribution of the paper is a strategy that enables a content-based recommender system to infer user interests by using machine learning techniques both on static content and tags. The main outcome of the experiments conducted is an improvement in the predictive accuracy of the tag-augmented recommender system compared to a pure content-based approach.**

I. Introduction

Cultural heritage personalization refers to supporting visitors in the selection and filtering of preferred artifacts and their corresponding descriptions, and in the creation of personalized tours. For example, PEACH (Personal Experience with Active Cultural Heritage) [1] is a research project for intelligent information presentation in museums, which aims to build an active, multimedia visitor guide, with strong personalization of all the information provided, so as to ensure that visitors, by expressing their affective attitude, are allowed to accommodate the museum tour according to their own interests and pace.

Because recommender systems have proved to be useful in helping users access to desired information (especially in domains where they are not expert or familiar with), they have found their way also in the context of museums, to support visitors in fulfilling a personalized experience and tour when visiting artworks collections. For instance, the CHIP project (Cultural Heritage Information Personalization) [2] is a research effort for enhancing personalized access to the collections of the Rijksmuseum in Amsterdam. CHIP combines Semantic Web technologies and content-based algorithms for deducing visitors' preference from a set of scored artifacts and then, recommending other artworks and related content topics. In particular, the recommendations of artworks are based on three properties, namely author, genre, and period.

When providing recommendations in cultural heritage context, information about collections must be taken into account because it can be as important as the artifacts themselves. Furthermore, the recent Web 2.0 (r)evolution has radically changed the role of people from passive consumers of information to that of active contributors who create and share new content. One of the forms of user-generated content (UGC) that has drawn more attention from the research community is *tagging*, which is the act of annotating resources of interests with free keywords, called *tags*, thus building a socially-constructed classification schema, called a *folksonomy* (folks + taxonomy). The Steve.museum consortium [3] has begun to explore the use of social tagging and folksonomy in cultural heritage personalization scenario, to increase audiences engagement with museums' collections. Supporting social tagging of artifacts and providing access based on the resulting folksonomy open museum collections to new interpretations, which reflect visitors' perspectives rather than curators' ones, and helps to bridge the gap between the professional language of the curator and the popular language of the museum visitor. Preliminary explorations conducted at the Metropolitan Museum of Art of New York have shown that professional perspectives differ significantly from those of naïve visitors. Hence, if tags are associated to artworks, the resulting folksonomy can be used as a different and valuable source of information to be carefully taken into account when providing recommendations to museum visitors. The goal of the paper can be formulated in form of a research question as follows:
In the context of cultural heritage personalization, does the integration of UGC (i.e., tags) cause an increase of the prediction accuracy in the process of recommending artifacts to users?

Content-based recommender systems analyze a set of documents, previously rated by an individual user, and learn a model or profile of user interests based on the *features* of the documents rated by that user. The profile is exploited to recommend new relevant items. This paper presents an approach in which the process of learning user profiles is performed both on static content and UGC. This research was conducted within the CHAT project (Cultural Heritage fruition & e-learning applications of new Advanced multimodal Technologies), that aims at developing new systems and services for multimodal fruition of cultural heritage content. We gathered data from the collections of the Vatican picture-gallery, for which both images and detailed textual information of paintings were available, and letting users involved in the study both rate and annotate them with tags.

The paper is structured as follows. Section II provides a description of our recommender system and how it handles users' tagging activity when building user profiles. Section III provides the description of the experimental session carried out to evaluate the proposed idea, and a discussion of the main findings. Finally, Section IV draws conclusions and provides directions for future work.

II. A Content-based Recommender System handling Tags

The inceptive idea behind this paper is to include folksonomies in ITR [4], a content-based recommender system developed at the University of Bari, by integrating *static* content describing the artworks of the collection with *dynamic* user-generated content. Tags are collected during the step, by letting users: 1) express their preferences for items by entering a numerical rating and 2) annotate rated items with free tags. FIRSt (Folksonomy-based Item Recommender SysTem) extends original ITR integrating user-generated content management.

The recommendation process is performed in three steps, each handled by a separate component. First, given a collection of documents, a preprocessing step is performed by the *Content Analyzer*, which uses the WordNet lexical database to perform *Word Sense Disambiguation* (WSD) on both static and dynamic content to identify correct senses, corresponding to concepts identified from words in the text. Then, a learning step is performed by the *Profile Learner* on the training set of documents, to generate a probabilistic model of the user interests. This model is the personal profile including those concepts that turn out to be most indicative of the user's preferences. Finally, the *Recommender* component implements a naïve Bayes text categorization algorithm, which is able to classify new documents as interesting or not for a specific user by exploiting the probabilistic model learned from training examples.

A. Content Analyzer: Semantic Indexing of Static and Dynamic Content

Here we describe the document representation technique used to build *semantic* user profiles based on the senses (meanings) of words found in the training documents. There are two crucial issues to address: First, a repository for word senses has to be identified; second, any implementation of sense-based document representation must solve the problem that, although words occur in a document, meanings do not, since they are often hidden in the context. Therefore, a procedure is needed for assigning senses to words: The task of WSD consists in determining which sense of an ambiguous word is invoked in a particular use of the word [5]. As for the sense repository, we adopted WordNet version 2.0. The basic building block for WordNet is the synset (SYNonym SET), a structure containing sets of words with synonymous meanings, which represents a specific meaning of a word. Our WSD algorithm, called JIGSAW, takes as input a document $d = [w_1, w_2, \ldots, w_h]$ encoded as a list of words in order of their appearance, and returns a list of WordNet synsets $X = [s_1, s_2, \ldots, s_k]$ $(k \leq h)$, in which each element s_j is obtained by disambiguating the *target word* w_i based on the *semantic similarity* of w_i with the words in its context. Notice that $k \leq h$ because some words, such as proper names, might not be found in WordNet, or because of bigram recognition. Semantic similarity computes the relatedness of two words. We adopted the Leacock-Chodorow measure, which is based on the length of the path between concepts in a IS-A hierarchy. Since WSD in not the focus of the paper, we do not provide here the complete description of the strategy adopted. More details are reported in [6]. What we would like to point out here is that the WSD procedure allows to obtain a synset-based vector space representation, called bag-of-synsets (BOS), that is an extension of the classical bag-of-words (BOW) model. In the BOS model a synset vector, rather than a word vector, corresponds to a document. FIRSt is capable of providing recommendations for items in any domain (e.g., films, music, books), as long as item properties can be represented in form of *textual slots*. Hence, in the context of cultural heritage personalization, an artwork can be generally represented by at least three slots, namely *artist*, *title*, and *description*. Besides, if museum visitors have a digital support to annotate artifacts, tags can be easily stored in a fourth slot, say *tags*, which is not static as the other three slots because tags evolve over time. In systems supporting social tagging, the number of tags used to annotate a given resource tend to grow initially, and then to decrease because users tend to reuse existing tags, especially the most common ones. This phenomenon is known as tag convergence. However, being free annotations, tags also tend to suffer from syntactic problems, like polysemy and synonymy, which hinder tag convergence. One way to cope with such a problem is to apply WSD to tags as well. This process allows the document representation model to evolve from using tags as mere keywords or strings, to using *semantic tags* and, consequently, *semantic folksonomies* of concepts. The text in each slot is represented by the BOS model by counting separately the occurrences of a synset in the slots in which it appears. More formally, assume that we have a collection of N documents. Let *m* be the index of the slot, for $n = 1, 2, ..., N$, the n-th document is reduced to four bag of synsets, one for each slot:

$$d_n^m = \langle t_{n1}^m, t_{n2}^m, \ldots, t_{nD_{nm}}^m \rangle$$

where t_{nk}^m is the k-th synset in slot s_m of document d_n and D_{nm} is the total number of synsets appearing in the m-th slot of document d_n. For all n, k and m, $t_{nk}^m \in V_m$, which is the vocabulary for the slot s_m (the set of all different synsets found in slot s_m). Document d_n is finally represented in the vector space by four synset-frequency vectors:

$$f_n^m = \langle w_{n1}^m, w_{n2}^m, \ldots, w_{nD_{nm}}^m \rangle$$

where w_{nk}^m is the weight of the synset t_k in the slot s_m of document d_n and can be computed in different ways: it can be simply the number of times synset t_k appears in slot s_m or a more complex TF-IDF score. All the text operations performed on documents are provided by a NLP tool developed at University of Bari, called META [7]. Our idea is that BOS-indexed documents can be used in a content-based information filtering scenario for learning accurate, *sense-based* user profiles, as discussed in the following section.

B. Profile Learner: Learning User Profiles from Static Content and UGC

We consider the problem of learning user profiles as a binary Text Categorization task since each document has to be classified as interesting or not with respect to the user preferences. Therefore, the set of categories is restricted to c_+, that represents the positive class (user-likes), and c_- the negative one (user-dislikes). The induced probabilistic model is used to estimate the *a posteriori* probability, $P(c_j|d_i)$, of document d_i belonging to class c_j. The algorithm adopted for inferring user profiles is a Naïve Bayes text learning approach, widely used in content-based recommenders, which is not described here due to space limitations. What we would like to point out here is that the final outcome of the learning process is a probabilistic model used to classify a new document in the class c_+ or c_-. Given a new document d_j, the model computes the a-posteriori classification scores $P(c_+|d_j)$ and $P(c_-|d_j)$ by using probabilities of synsets contained in the user profile and estimated in the training step. The profile contains the user identifier and the *a-priori* probabilities of liking or disliking an item, apart from its content. Moreover, the profile is structured in two main parts: *profile_like* contains features describing the concepts able to deem items relevant, while features in *profile_dislike* should help in filtering out not relevant items. Each part of the profile is structured in four slots, resembling the same representation strategy adopted for artworks. Each slot reports the features (WORDNET identifiers) occurring in the training examples, with corresponding frequencies computed in the training step. Frequencies are used by the Bayesian learning algorithm to induce the classification model (i.e. the user profile) exploited to suggest relevant artworks in the recommendation phase.

III. EXPERIMENTAL EVALUATION

The goal of the experimental evaluation was to compare the predictive accuracy of our recommender system when 1) user profiles are learned from static content only; 2) both static content and UGC are used in the learning process. In addition, to properly investigate the effects of including social tagging in the recommendation process, a distinction has to be made between considering, for an artifact rated as interesting by a user, either the whole folksonomy (i.e., the community tags used by all visitors to annotate that artifact), or only the tags entered by that user for that artifact (i.e., the user's contribution to the whole artifact folksonomy). For this purpose, we designed several experiments, described in the following.

A. Users and Dataset

The dataset considered for the experiments is represented by 45 paintings chosen from the collection of the Vatican picture-gallery. The dataset was collected using screenscraping bots, which captured the required information from the official website[1] of the Vatican picture-gallery. In particular, for each element in the dataset an image of the artifact was collected, along with three textual properties, namely its title, artist, and description. We involved 30 non-expert users (average age $\sim$ 25) who volunteered took part in the experiments. Users were requested to interact with a web application (Figure 1), in order to express their preferences for all the 45 paintings in the collection. The preference was expressed as a numerical vote on a 5-point scale (1=strongly dislike, 5=strongly like). Moreover, users were left free to annotate the paintings with as many tags as wished. For the overall 45 paintings in the dataset, 4300 tags were used.

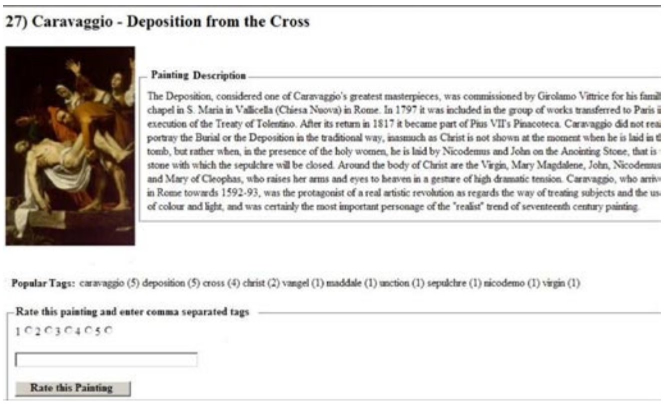


Fig. 1. Gathering user ratings and tags

B. Design of the Experiment and Evaluation Metrics

Since FIRSt is conceived as a text classifier, its effectiveness can be evaluated by classification accuracy measures, namely *Precision* and *Recall*. Precision (Pr) is defined as the number of relevant selected items divided by the number of selected items. Recall (Re) is defined as the number of relevant selected items divided by the total number of relevant items available. We adopted these specific measures because we are interested in measuring how *relevant* a set of recommendations is for a user. In the experiment, a painting is considered as *relevant* by a user, if the rating is greater than or equal to 4, while FIRSt considers a painting as relevant if the a-posteriori probability of class likes is greater than 0.5. We designed 5 different experiments, depending on the type of content used for training the system:

- EXP #1: STATIC CONTENT - only title, artist and description of the painting, as collected from the official website of the Vatican picture-gallery
- EXP #2: PERSONAL TAGS - only tags provided by a user on a painting
- EXP #3: SOCIAL TAGS - all the tags provided by all the users on a painting
- EXP #4: STATIC CONTENT + PERSONAL TAGS
- EXP #5: STATIC CONTENT + SOCIAL TAGS

[1] `http://mv.vatican.va/3_EN/pages/PIN/PIN_Main.html`

All experiments were carried out using the same methodology, consisting in performing one run for each user, scheduled as follows:

1) select the appropriate content depending on the experiment being executed;
2) split the selected data into a training set *Tr* and a test set *Ts*;
3) use *Tr* for learning the corresponding user profile;
4) evaluate the predictive accuracy of the induced profile on *Ts*.

The methodology adopted for obtaining *Tr* and *Ts* was the K-fold cross validation [8], with $K = 5$. Given the size of the dataset (45), applying a 5-fold cross validation technique means that the dataset is divided into 5 disjoint partitions, each containing 9 paintings. The learning of profiles and the test of predictions were performed in 5 steps. At each step, 4 (K-1) partitions were used as the training set *Tr*, whereas the remaining partition was used as the test set *Ts*. The steps were repeated until each of the 5 disjoint partitions was used as the *Ts*. Results were averaged over the 5 runs.

C. Discussion

Results of the 5 experiments are reported in Table I, averaged over the 30 users.

TABLE I
RESULTS OF THE K-FOLD CROSS VALIDATION

Type of content	Precision	Recall	F1
Exp #1: Static Content	75.86	94.27	84.28
Exp #2: Personal Tags	75.96	92.65	83.26
Exp #3: Social Tags	75.59	90.50	82.13
Exp #4: Static Content + Personal Tags	78.04	93.60	84.93
Exp #5: Static Content + Social Tags	78.01	93.19	84.73

The main finding is that the integration of UGC (whether social or personal tags) causes an increase of precision in the process of recommending artifacts to users. More specifically, precision of profiles learned from both static content and tags (hereafter, augmented profiles) outperformed the precision of profiles learned from either static content (hereafter, content-based profiles) or just tags (hereafter, tag-based profiles). The improvement ranges between 2% and 2.40%. Another interesting finding is that precision of content-based profiles is comparable with that of tag-based profiles. Although this result may suggest that just tags are sufficient for providing accurate recommendations, a decrease of recall (-1.62% with personal tags, -3.77% with social tags) actually shows that static content cannot be neglected even if tags are available. The higher decrease of recall registered with social tags leads to conclude that community tags introduce some noise in the recommendation process (relevant paintings are filtered out due to wrong advice by other users). The general conclusion of the comparison between content-based profiles and augmented profiles is that a significant increase of precision corresponds to a slight and physiological loss of recall. The overall accuracy of augmented profiles (F1 about 85%) is considered satisfactory.

IV. CONCLUSIONS AND FUTURE WORK

In this paper we have investigated how to effectively combine existing content-based filtering algorithms with UGC, in the context of cultural heritage personalization. The main contribution of the paper is an approach in which machine learning techniques are adopted to infer user profiles both from static content, as in classical content-based recommender, and UGC, namely tags provided by users to freely annotate artworks. The main outcome of the experiments performed to evaluate the proposed approach is that the integration of UGC causes an increase of precision in the process of recommending artifacts to users.

By definition, social tags used for annotating a painting include personal tags. However, the findings from the experiments with social tags ran counter our expectation because, as compared to the use of personal tags only, a decrease of precision and recall was observed. To gain more insights on the effects of community-generated content, we need to 1) perform an analysis of what tags are used to build the folksonomies and how they affect the user profile generation; 2) replicate the experiments with a more heterogeneous community, involving experts in the art domain so as to identify differences with the tagging activity of naïve users.

ACKNOWLEDGEMENTS

This research was funded by MIUR (Ministero Universitá e Ricerca) under the contract Fondo per le Agevolazioni alla Ricerca, D.L: 297/99 "CHAT - Cultural Heritage fruition & e-learning applications of new Advanced (multimodal) Technologies" (2006-2008).

REFERENCES

[1] O. Stock, M. Zancanaro, P. Busetta, C. Callaway, A. Krüger, M. Kruppa, T. Kuflik, E. Not, and C. Rocchi, "Adaptive, intelligent presentation of information for the museum visitor in peach," *User Modeling and User-Adapted Interaction*, vol. 17, no. 3, pp. 257–304, 2007.

[2] Y. Wang, L. Aroyo, N. Stash, and L. Rutledge, "Interactive user modeling for personalized access to museum collections: The rijksmuseum case study," in *User Modeling 2007, Proc. 11th Int'l Conf.*, ser. LNCS, C. Conati, K. F. McCoy, and G. Paliouras, Eds., vol. 4511. Springer, 2007, pp. 385–389.

[3] J. Trant and B. Wyman, "Investigating social tagging and folksonomy in art museums with steve.museum," in *Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland*, May 2006.

[4] M. Degemmis, P. Lops, and G. Semeraro, "A content-collaborative recommender that exploits wordnet-based user profiles for neighborhood formation," *User Model. User-Adapt. Interact.*, vol. 17, no. 3, pp. 217–255, 2007.

[5] C. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. Cambridge, US: The MIT Press, 1999, ch. 7: Word Sense Disambiguation, pp. 229–264.

[6] G. Semeraro, M. Degemmis, P. Lops, and P. Basile, "Combining learning and word sense disambiguation for intelligent user profiling." in *Proc. 20th Int'l Joint Conf. on Artificial Intelligence IJCAI-07*, 2007, pp. 2856–2861, m. Kaufmann, San Francisco, California. ISBN: 978-I-57735-298-3.

[7] P. Basile, M. de Gemmis, A. Gentile, L. Iaquinta, P. Lops, , and G. Semeraro, "META - MultilanguagE Text Analyzer," in *Proc. Language and Speech Technnology Conf. - LangTech 2008, Rome, Italy*, 2008, pp. 137–140.

[8] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *IJCAI*, 1995, pp. 1137–1145. [Online]. Available: citeseer.ist.psu.edu/kohavi95study.html

2nd VARIAZIONI Workshop on Technologies for Content Enrichment in the Web2.0 era

Organised by

VARIAZIONI EC e-Content Plus Project

http://www.variazioniproject.org/

AXMEDIS 2008: proceedings of the 4th International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution. Workshops, Panels and Industrial Applications. Florence, Italy 17-19 November 2008
edited by Paolo Nesi, Kia Ng, Jaime Delgado
ISBN: 978-88-8453-811-6 (online). ISBN 978-88-8453-810-9 (print).

The CACAO project: a multilingual interface to Library Catalogues

(Extended Abstract)

Alessio Bosca, Claude Roux
CELI
Xerox Research Centre Europe
{alessio.bosca@celi.it, claude.roux@xerox.xrce.com}

Introduction

Managing the development and delivery of multilingual electronic library services is one of the major current challenges for making digital content in Europe more accessible, usable and exploitable. Digital libraries and OPAC-based traditional libraries are the most important source of reliable information, daily used by scholars, researchers, knowledge workers and citizens to conduct their working (and leisure) activities. Facilitating access to multilingual document collections therefore is an important way of supporting the dissemination of knowledge and cultural content.

CACAO (Cross-language Access to Catalogues And On-line libraries) project proposes an innovative approach for accessing, understanding and navigating multilingual textual content in digital libraries and OPACs, enabling European users to better exploit the available European electronic content at their disposal.

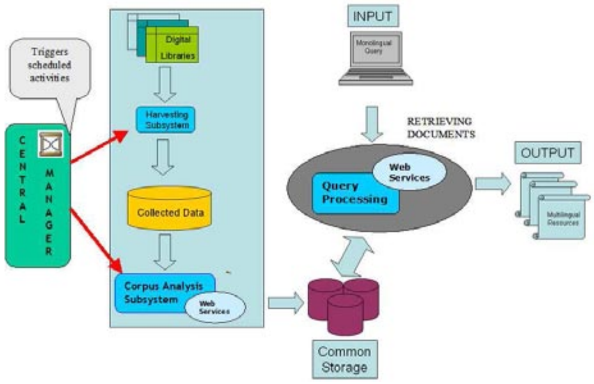
By coupling sound Natural Language Processing techniques with available information retrieval systems the project aims at the delivery of a non-intrusive infrastructure to be integrated with current OPAC and digital libraries. The result of such integration will be the possibility for the user to type in queries in his/her own language and retrieve volumes and documents in any available language.

CACAO aims at offering cross-lingual and cross-border access to the content of classical and digital libraries and enabling users to find digital content irrespective of the language. In fact, in a context of interlaced cross-border libraries, such as the one proposed by META OPAC, the absence of a cross-language perspective is likely to cause a substantial impasse: if a user wanted to access a META OPAC including the National Libraries of France, Germany, Italy, Poland and Hungary, s/he would have to type five queries in five different languages. Much of the advantage of having a unique access point is thus lost.

CACAO project proposes a system based on the assumptions that users look more and more at library contents using free keyword queries (as those used with a web search engine) rather than more traditional library-oriented access (e.g. via Subject Heading); therefore, the only way to face the cross-language issue is by translating the query into all languages covered by the library/collection (rather than, for instance, translating subject headings). The system will then yield results in all desired languages.

Validation is another important aspect in the project: all CACAO core technologies are indeed proven, but they have never been massively deployed in the field of digital libraries. CACAO aims at crossing the chasm between sound innovation and adoption by library institutions for real life purposes.

Architecture overview



CACAO proposes the development of an infrastructure for multilingual access to digital content, including an information retrieval system able to search for books and texts in all the available languages. The core of the search engine takes advantage of information contained in existing catalogues and texts of the digital libraries that is enriched by means of NLP techniques such as word sense disambiguation and named entities recognition. The goal of such integration is to avoid confusing the user by providing irrelevant results due

to bad translations and thus enabling a better access to the digital content.
The general architecture of the Cacao system could be summarized as the result of the interactions of few functional subsystems, coordinated by a central manager and reacting to external stimuli represented by end users queries:

- **Harvesting** subsystem is in charge of collecting data from digital libraries, abstracting from the multiplicity of standards and protocols, and storing them into a repository.
- **Corpus Analysis** subsystem performs specific analysis on the data collected from libraries and infers new information used to support query processing and resource retrieval (e.g. query expansion, terms disambiguation,..).
- **Web Services** subsystem represents third party software providing specific services (e.g. linguistic analysis, translations,..).
- **Query Processing** subsystem: a set of components is devoted to process the original monolingual user query, transforming and enriching it by means of translations and expansions.

Content Enrichment

CACAO approach to multilingual access is based on the integration of a standard IR engine with multilingual thesauri and multilingual lexicons. However a simple, direct integration would provide poor results since records of digital catalogues often contain only small portions of text and the noise brought in by the query translation layer further worsen the situation. Therefore any single fragment of text needs to be linguistically "enriched" in order to guarantee an optimal retrieval.
The strategy adopted by CACAO with respect to content enrichment aims at integrating the search indexes used by the IR system rather than the original records from the libraries; such enrichment is operated by adopting the following technologies:

1. Enrichment of the query via thesauri: the simple query "*plants*" could be enriched by synonyms/hyperonyms/hyponyms such as horticultural, seeds etc. This would allow books such as "*American Horticultural Society Encyclopaedia of Plants and Flowers*" by Christopher Brickell or "*From Seed to Plant*" by Gail Gibbons to receive more emphasis than "*The Parachute Plant*", a thriller by Bill Carrigan.
2. Enrichment of the query via corpus-based expansion lists: from the point of view of the user, this technology is the same as the previous one. The only difference is that such related terms are induced on the basis of the catalogue rather than being stored into a static repository.
3. Tagging of the text in DB records by using a part of speech tagger (i.e. disambiguating the syntactic category of words). As simple as it might seem, this enrichment will allow the system to avoid retrieving a title such as "*Plant Them Deep*" by Aimee Thurlo, David Thurlo with a query such as *plant*.

Improving Translations

Within CACAO project another aspect where contents enrichment has a strong impact is the improvement of linguistic resources and in particular translation dictionaries. CACAO system is based on translation dictionaries; however, there is probably no single translation dictionary that would be able to cover all digital content either in a library catalogue or in the texts of a digital library.
A first strategy to be adopted in order to compensate for possible lack of translation coverage is query expansion.
The second, probably more innovative approach is based on user input. An analysis of the web logs of a university library, shows that about 40% of the queries are "duplicated" in at least two languages. Indeed, if we could store the translations implicitly provided by the user, we could add items which are I) relevant to users; II) reflecting users' perception of the translation of a given word in a different language.
From a technical point of view, this approach raises some major challenges. Indeed, the fact that two queries issued by the same user are temporally adjacent is not necessarily proof that the second is a translation of the first. Therefore, it is important to set up methodologies to isolate possible translation pairs. In order to detect these cases CACAO system exploits a method based on semantic web vectors. The basic idea is to gather from the web (via queries to search engines) a set of documents strongly related to the original term (**st**).
By using standard NLP technologies, these documents are analyzed, and the terminological items are extracted (let **ST** be these terms). The same operation is performed on the candidate translation (**tt**), thus generating a set **TT** of words in the target language. **ST** is then translated, using the available resources, into a set of translated target words (**STT**). By measuring the intersection between **STT** and **TT**, the system will be able to predict the likeliness of **tt** to be a translation of **st**.

Content Enrichment in the GAMA Project

A. Lüdtke*, B. Gottfried*, O. Herzog*, G. Ioannidis†, M. Leszczuk‡, and V. Šimko§

* Technologie-Zentrum Informatik, Universität Bremen
Am Fallturm 1, D-28359 Bremen, Germany

† IN2 Search Interfaces Development Ltd.
Fahrenheitstrasse 1, D-28359 Bremen, Germany

‡ Department of Telecommunications, AGH University of Science and Technology
al. Mickiewicza 30, PL-30059 Krakow, Poland

§ CIANT, International Centre for Art and New Technologies
Kubelíkova 27, CZ-13000 Prague 3, Czech Republic

***Abstract*—With its eContentplus programme, the European Community supports projects that help develop the "i2010: Digital Libraries" initiative. The project GAMA (Gateway to Archives of Media Art) is amongst those projects to participate in this endeavour with a Community funding of 1.2 million Euro. The GAMA consortium comprises 19 institutions from Europe's culture, art, and technology sectors from 17 European countries. The objective is to establish a portal for online access to Europe's most important digital archives and libraries on media art. This paper describes enrichment of media content (mainly video art) with metadata and handling of metadata within the GAMA architecture.**

I. Introduction

Through the GAMA portal media art content - mainly video art - is made accessible for the interested public, e.g., for curators, artists, academics, researchers, and mediators. In the GAMA portal metadata plays an important role in advanced search, that includes *Query by Example* (*QbE* hereafter) subsystems for video and still image content, and in browsing interfaces that allow for browsing content by criteria like, e.g., artist, genre or production year.

The idea behind the GAMA architecture is that, while provision of full, high-quality content is still up to the content providers, the GAMA system

1) collects metadata from the content providers' local databases in a centralized repository,
2) performs content-based analysis of media content and produces audiovisual descriptions,

and provides centralized advanced search and browsing functionality based on that in the GAMA portal for all connected archives.

II. Media content and metadata

Media content is enriched with metadata mainly in two ways. One is metadata directly imported from the content providers' databases. The challenge here is to integrate data from different sources with heterogeneous data models and ensure its interoperability. The other is content-based metadata extracted from the raw media content, e.g., descriptions of audiovisual characteristics of media content. This is typically not available from the content providers and extracted by the central content-based indexing service. Figure 1 displays relevant parts of the GAMA architecture.

III. GAMA system architecture

The GAMA system architecture highly aims at simplicity for content providers as these cannot be assumed to be able to host a complicated infrastructure in general. One of the major goals in design of the GAMA system architecture was to minimize the required infrastructure at the archives. So, in the simplest case, it would not even be necessary for content providers to have a permanent internet connection or to give direct access to their media or database servers. In this case it is also possible to also upload media and metadata (e.g., database dumps) to the GAMA servers. This is important as part of the content providers does not allow direct access to their servers for legal reasons. Furthermore, no software components need to run at the side of the content providers at all. The GAMA system architecture is organized in services running in distributed locations. Services relevant in the context of this paper - this is services dealing with metadata - are the content-based indexing service and the central metadata repository.

The central metadata repository is the central storage component for all metadata in the GAMA system. It provides a SOAP interface for data ingest and querying the repository. Metadata available from the archives (e.g., database dumps) is either automatically downloaded from the archives servers or uploaded by content providers and mapped by database adaptors.

The content-based indexing service analyses raw media and generates content-based descriptions. For an overview of analysis modules refer to section V. Additionaly there are two *QbE* subsystems based on the extracted features (see section VI). Media is either automatically downloaded from

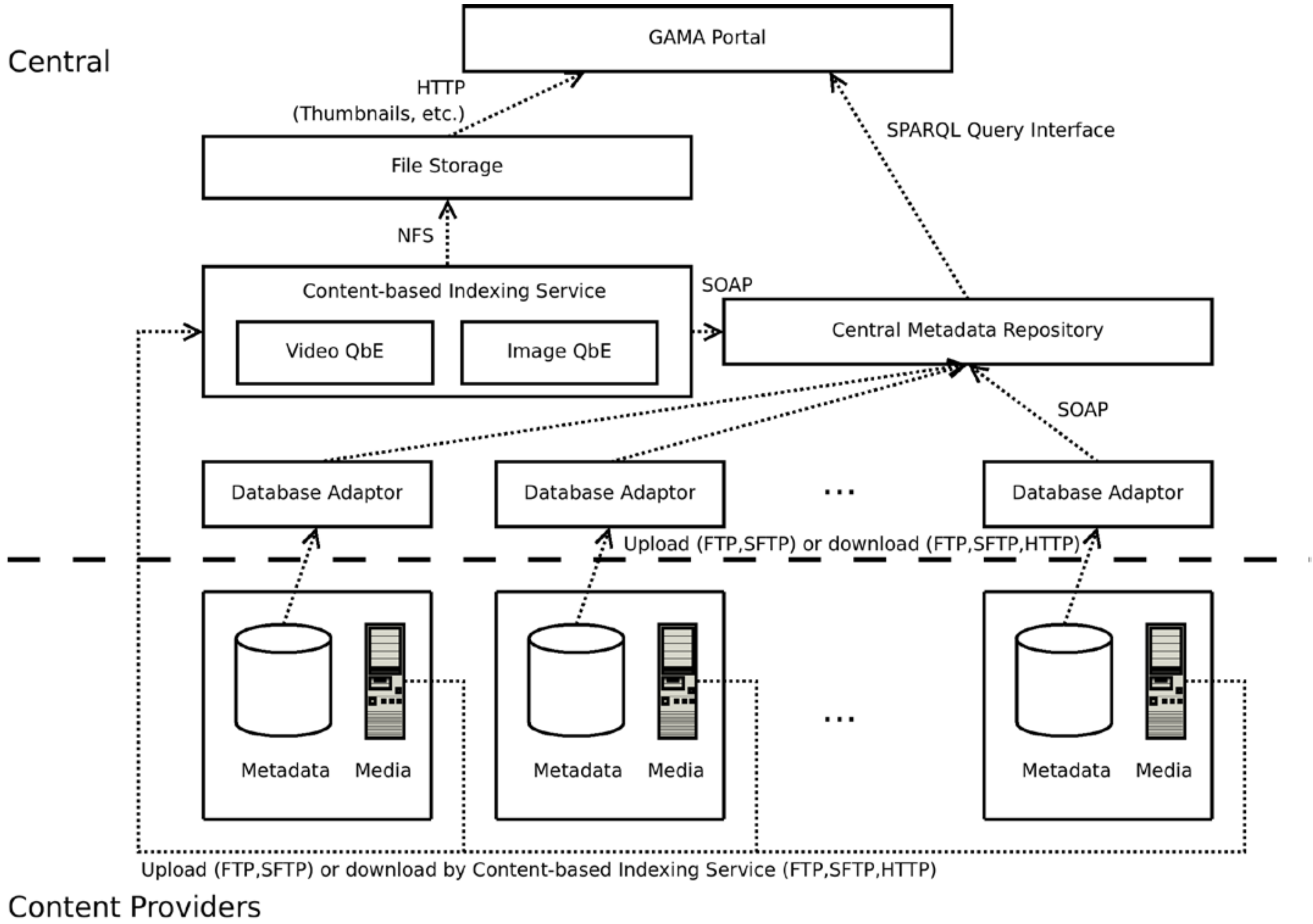


Fig. 1. Sketch of relevant parts of the GAMA system architecture.

the archives servers or uploaded by content providers. Metadata generated by the content-based indexing service (e.g., audiovisual descriptions) is exported to the central metadata repository through the SOAP interface for data ingest. Part of the output is file-based, e.g., thumbnails of keyframes that are extracted per video shot. File-based output is written directly to the file storage.

Both, metadata generated by the content-based indexing service and metadata imported from the archives, is then accessed by the GAMA portal through the query interface of the central metadata repository (see section VI). File-based outputs of content-based analysis (e.g., thumbnails of keyframes) are accessible for use within the portal through a web server via HTTP.

IV. Metadata import from content providers' local databases

Metadata from content providers' local databases is imported into the central repository through so-called *database adaptors* (see Figure 1). The data model of the central repository is based on RDF[1], which is a flexible solution with regard to heterogeneous data models of the archives. Database adaptor implementations are content provider specific and map data from local databases to RDF/XML according to the GAMA RDF schema, and transfer it to the central GAMA metadata repository.

Metadata from content providers' databases typically includes information on artworks (e.g., title, date of creation), manifestations of artworks (e.g., format, length), persons such as artists or curators (e.g., name, date of birth), archives (e.g., name, homepage), and similar and relations thereof, such as "an *artwork* is *provided by* an *archive*" or "a *person* (artist in that case) is *author of* an *artwork*".

V. Content-based analysis and extraction of audiovisual descriptions

Content-based metadata is extracted by the content-based indexing service which has a module-based structure. Several modules extract content-based features, e.g., *Automatic Speech Recognition* (*ASR* hereafter), still image and video *Optical*

[1]Resource Description Framework, see `http://www.w3.org/RDF`

Character Recognition (*OCR* hereafter), face recognition, shot boundary detection, as well as extraction of a set of MPEG-7 visual, audio, and video-specific descriptors is applied. Part of the extracted content-based metadata is textual data that directly allows for text-based querying, e.g., results of *ASR* and *OCR*. Other metadata is utilized within the two *QbE* subsystems for image and video content. Additionally, e.g., keyframes are extracted from a shot and used as elements to visually enhance the GAMA user interface.

Feature extraction modules applied within the content-based analysis of the GAMA system are described in the following sections.

A. Shot boundary detection and keyframe extraction

The shot boundary detection and key frame extraction module based on the approach described in [1] extracts shot boundaries within input video content and thereby extracts representative frames per shot. Results have the form of start and end frame number per shot and associated images (key frames). Key frames are extracted in three resolutions:

- Original resolution of input video
- Large image (fixed-size for use in the portal)
- Thumbnail (fixed-size for use in the portal)

Extracted shots are the basis for video-based *QbE* within the GAMA system. Video features are extracted and matched on a shot basis. Queries for video *QbE* are shots for which associated key frames are displayed in the GAMA portal and linked with *QbE* functionality. Extracted key frames in original resolution are fed into the analysis process for feature extraction and used as a part of video *QbE*.

Key frames can also be used in various ways as visual elements within the GAMA portal, e.g., in result lists or in a detailed view of a video as an overview of the temporal structure.

B. MPEG-7 audiovisual descriptors

The MPEG-7 (“Multimedia content description interface”) standard has been selected for describing the audiovisual content. The descriptions are generated as a result of analysis in both the visual and audio domain.

For MPEG-7 based visual [2] indexing a subset of MPEG-7 visual descriptors was chosen. MPEG-7 visual descriptors can further be divided into descriptors applicable for still images/video frames and descriptors that are extracted on the basis of video segments (shots in case of GAMA). For still images/video frames the following MPEG-7 visual descriptors are extracted:

- Color Layout Descriptor
- Dominant Color Descriptor
- Scalable Color Descriptor
- Color Structure Descriptor
- Edge Histogram Descriptor

These descriptors are utilized for both, still-image and video matching [3] within the respective *QbE* subsystems (see Figure 2 with an example utilizing images downloaded from Flickr[2]). Within the video *QbE* subsystem these descriptors are extracted from key frames extracted by the shot boundary detection module (see last paragraph).

As the goal of the MPEG-7-based *QbE* subsystem is to provide the best matches to the query object, currently some user tests are executed, which aim at selecting the best combination of the above-mentioned visual descriptors. The user tests are tailored to the media art content and consumers. The tests will allow for constructing the optimal model for combining the MPEG-7 visual descriptors.

Additionally, for video *QbE* there are video-specific visual and audio descriptors extracted on a shot basis. Two video-specific visual descriptors are extracted per video shot:

- Motion Activity Descriptor
- Camera Motion Descriptor

And finally two audio [4] descriptors are extracted per video shot. These are:

- Audio Spectrum Centroid Descriptor
- Audio Power Descriptor

MPEG-7 audiovisual descriptions are utilized within the *QbE* subsystems for still image (where applicable) and video *QbE*. For comparison of descriptors the distance measures proposed by the MPEG-7 standard [3] are applied.

C. TZI PictureFinder

TZI PictureFinder [5] is an extremely fast matching solution for image-to-image (or frame-to-frame) matching based on the distribution of visual features such as color and texture. It is especially optimized for fast matching within large datasets. In the context of the GAMA system TZI PictureFinder is used as a fast pre-filtering solution within both *QbE* subsystems to considerably improve the performance of the generation of pre-calculated result lists that are then exported to the central RDF repository (see section VI).

*D. Optical Character Recognition (*OCR*)*

The GAMA *OCR* module extracts text from video frames and still images. For input videos *OCR* is applied on every n-th video frame. *OCR* results are further filtered to avoid misdetections. In the GAMA system *Tesseract OCR*[3], a freely available open source optical character recognition engine, is used as *OCR* software. Results are words occurring in still images or video frames and time of occurrence (on a shot basis) in case of videos. The *OCR* module produces textual output that directly allows for text-based querying in the GAMA portal.

*E. Automatic Speech Recognition (*ASR*)*

The GAMA *ASR* module extracts spoken text from audio tracks of input videos. In the GAMA system the Microsoft *Speech Application Programming Interface* (SAPI)[4] is used as *ASR* software. Results are spoken words and time of

[2]See `http://www.flickr.com/`

[3]See `http://code.google.com/p/tesseract-ocr`

[4]See `http://www.microsoft.com/speech/speech2007/default.mspx`

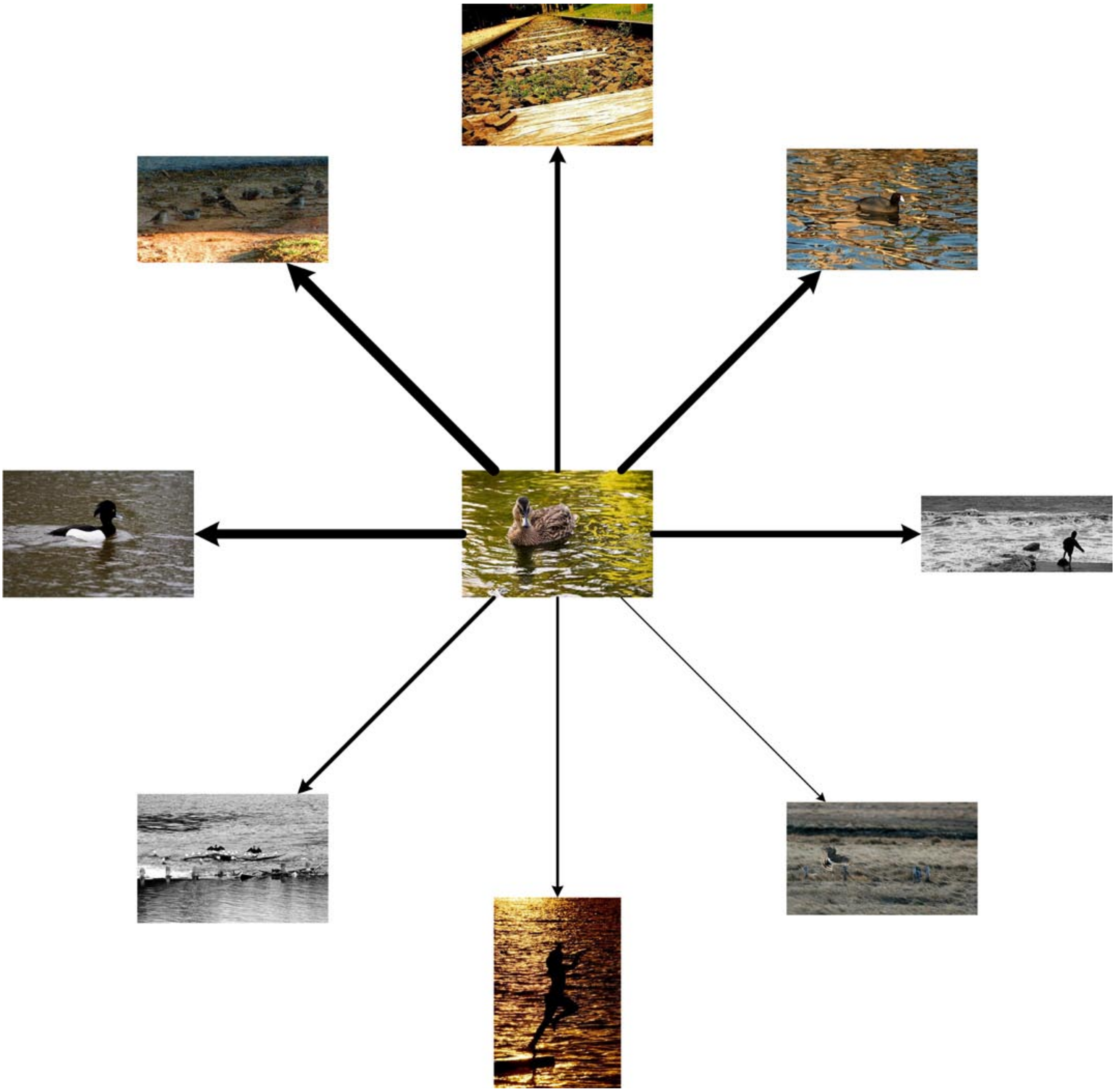


Fig. 2. An example of an image *QbE* system. The query image is in the middle whilst the best 8 matches (according to the MPEG-7 Edge Histogram Descriptor) are located around. Arrow widths indicate similarity (the wider arrow, the higher similarity).

occurrence (on a shot basis). The *ASR* module produces textual output that directly allows for text-based querying in the GAMA portal.

*F. Face Recognition (*FR*)*

The GAMA *FR* module detects faces occurring in input videos and recognizes faces occurring multiply. In the GAMA system *CMU Face Detector* [6] is used for detecting faces whilst *CSU Face Identification Evaluation System* [7] provides standard face recognition algorithms and standard statistical methods for comparing face recognition algorithms. Results are identifiers of actors and time of occurrence in a video. This enables searching for occurrences of the same person (actor) in other videos or other parts of a video for a given actor, which is part of advanced search functionality in the GAMA portal.

G. Sound event detection and music/speaker audio segmentation

The sound event detection module based on [8] is a generic component that can be trained to detect certain predefined audio events based on low-level audio features and classification through application of Support Vector Machines. In the context of GAMA it will be trained for a number of events that are currently discussed with content partners. It enables search for video segments where certain audio events occur.

In the context of the SVP project[5] [9] at TZI it has been successfully applied to identify video segments with spoken words or background music. These models will be applied within GAMA to realize a music/speaker audio segmentation. The speaker/music audio segmentation finds segments with spoken text and/or background music in the audio tracks of input videos. It will be used as a feature within video *QbE*.

VI. Querying the RDF repository and QbE subsystems

The common query interface for both, metadata from the archives' databases collected by database adaptors and automatically extracted metadata from content-based indexing modules and especially *QbE* for images and videos, is the SPARQL[6] query interface of the central RDF repository. For efficiency reasons, with respect to response times and expected traffic on the GAMA portal, also *QbE* operations are not performed online. Instead, ordered lists of a fixed number of results are pre-computed and stored within the RDF repository for every potential query, (every shot of a video indexed by the GAMA content-based indexing service) and updated on a regular basis. This is applied for both image as well as video *QbE*.

An important advantage of this approach is that *QbE* results can directly be aligned with other search criteria through the same query interface (e.g., search for artworks of a specific artist or from a certain category). All queries are formulated in SPARQL, which makes the query interface more flexible, efficient, and also much more consistent.

The GAMA portal makes use of this in various ways, e.g., through text fields for keyword search, filtering of result lists by certain criteria such as genre or artist, or by application of *QbE* aligned with other criteria. Browsing interfaces within the GAMA portal allow for browsing media content based on available metadata, e.g., in list interfaces.

The following sections describe the approach for result list generation in the video and image *QbE* subsystems.

A. Video QbE *subsystem*

The approach applied for video *QbE* within the GAMA system is shot-based. Features (or descriptors in the sense of MPEG-7) are extracted and matched on the basis of video shots (see section V-A). A shot-based approach was chosen, because, while for a complete video audiovisual characteristics might severely change over time, within a single shot these characteristics are much better defined in general.

Subsequently, also queries within the video *QbE* subsystem are single shots. The result of a query is a list of videos containing similar shots. As distance per video the distance of the best matching shot can be assummed. As an alternative, a voting approach based on all matching shots per video in the best N matches is currently evaluated. Result lists displayed to the user within the GAMA system are video-based, so this approach is consistent with the common approach for result presentation in the GAMA portal and video *QbE* can be combined with other search critera such as search for vides by genre, artist or similar in this manner.

The core functionality of the video *QbE* subsystem of the content-based indexing service is pre-generation of result lists for every potential query, which is every indexed shot in the sense of the GAMA system. Result lists are computed on the basis of all features (or descriptors) and distance measures (see section V). For a description of the approach for combination of distance measures refer to section VI-C. These lists are then exported to the central RDF repository for each shot.

B. Image QbE *subsystem*

The core functionality of the image *QbE* subsystem of the content-based indexing service, similar to the video *QbE* subsystem, is pre-generation of result lists per image in the database. Result lists are computed on the basis of all features (or descriptors) introduced in section V applicable for still images. For a description of the approach for combination of distance measures refer to section VI-C. These lists are then exported to the central RDF repository per image.

C. Pre-filtering and combination of feature distances

Both *QbE* subsystems for image and video matching are based on multiple features (descriptors in the sense of MPEG-7) and feature distances, and both pre-generate result lists for every potential query (images for image *QbE* and a shots for video *QbE*). Please note that especially for MPEG-7 descriptors a search according to a single descriptor typically requires a one-to-one comparison of the query descriptor with all database descriptors.

The worst-case estimation for database size in GAMA is one million shots for the video *QbE* subsystem and one million images for the image *QbE* subsystem overall, so an exhaustive search for all descriptors over the complete dataset is not feasible. To overcome this, *TZI PictureFinder* [5] (see section V-C) is applied as a fast pre-filtering solution within both, image and video *QbE*.

First, a (large) ordered subset of the N_1 (assume N_1 fixed, e.g., $N_1 = 5000$) best matching images or shots (here according to visual similarity of key frames extracted by the shot boundary detection and keyframe extraction module, see section V-A) is selected through a query to the PictureFinder system. Using PictureFinder this ordered subset can be computed in approx. 250ms on standard hardware for a database of one million images/shots. For all images/shots in

[5]See http://www.tzi.de/svp

[6]SPARQL Protocol and RDF Query Language, see http://www.w3.org/TR/rdf-sparql-query

this set a weighted combination of normalized distances for all features is then computed and the set is reordered according to this distance. The best matching N_2 (assume N_2 fixed, e.g., $N2 = 100$) images/shots then form the ordered set of results for a query.

VII. Conclusion

This paper described the approach for the enrichment of media content with metadata applied in the context of the GAMA project. The focus of this paper was to describe all components dealing with metadata in the GAMA architecture, which can be assumed as the "GAMA metadata engine". This covers the centralized accumulation of metadata through database adaptors as well as the content-based analysis of media content by a central service. As has been shown, the centralized accumulation of available metadata from the archives and further enrichment of media content with audiovisual descriptions from content-based indexing will significantly improve the searchability of the underlying content.

Acknowledgment

The GAMA project is funded under the eContentplus programme, a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable. The authors further would like to thank the whole GAMA consortium.

References

[1] A. Miene, A. Dammeyer, T. Hermes, and O. Herzog, "Advanced and Adapted Shot Boundary Detection," in *Proc. of ECDL WS Generalized Documents*, D. W. Fellner, N. Fuhr, and I. Witten, Eds., 2001, pp. 39–43.

[2] MPEG, "Multimedia content description interface - Part 3: Visual," International Organization for Standardization, ISO/IEC JTC 1/SC 29/WG 11, Coding of Moving Pictures and Audio, Tech. Rep. ISO/IEC 15938-3:2002, 2002.

[3] ——, "Multimedia content description interface - Part 8: Extraction and use of MPEG-7 descriptors," International Organization for Standardization, ISO/IEC JTC 1/SC 29/WG 11, Coding of Moving Pictures and Audio, Tech. Rep. ISO/IEC 15938-8:2002, 2002.

[4] ——, "Multimedia content description interface - Part 4: Audio," International Organization for Standardization, ISO/IEC JTC 1/SC 29/WG 11, Coding of Moving Pictures and Audio, Tech. Rep. ISO/IEC 15938-4:2002, 2002.

[5] T. Hermes, A. Miene, and O. Herzog, "Graphical Search for Images by PictureFinder," *Multimedia Tools and Applications. Special Issue on Multimedia Retrieval Algorithmics.*, vol. 27, no. 2, pp. 229–250, November 2005.

[6] H. A. Rowley, "CMU Face Detector," Carnegie Mellon University, 2007, http://www.cs.cmu.edu/ har/faces.html.

[7] D. Bolme, R. Beveridge, M. Teixeira, and B. Draper, "The CSU Face Identification Evaluation System: Its Purpose, Features and Structure," in *International Conference on Vision Systems (ICVS)*. Graz, Austria: Springer-Verlag, Apr 2003, pp. 304–311.

[8] D. Möhlmann, "Sound Detection and Recognition using Progressive Feature Extraction," Master's thesis, Universität Bremen, 2007.

[9] C. Brachmann, H. I. Chunpir, S. Gennies, B. Haller, T. Hermes, O. Herzog, A. Jacobs, P. Kehl, A. P. Mochtarram, D. Möhlmann, C. Schrumpf, C. Schultz, B. Stolper, and B. Walther-Franks, "Automatic Generation of Movie Trailers using Ontologies," *IMAGE - Journal of Interdisciplinary Image Science*, vol. 5, pp. 117–139, 2007.

Metadata and Licence Model for Music Resources in VARIAZIONI

Carlos A. Iglesias
Germinus XXI (Grupo Gesfor)
cif@germinus.com

Víctor Torres
Univ. Pompeu Fabra
victor.torres@upf.edu

Francesco Spadoni
Rigel Engineering
spadoni@rigel.it

Abstract

This article introduces the music metadata model and the licence model defined within the eContentPlus VARIAZIONI project[1], based on FRBR. After analysing the limitation of traditional cataloguing approaches for music, and the difficulties of applying FRBR, the Variazioni metadata model defines a flexible model that takes into account the different nature of musical assets (libretto, master class, live recording, poster, etc.) as well as the musical analyst requirements and structural metadata between different media files. This metadata model is complemented by a licence model defined in MPEG-21 and implemented with Axmedis technology.

1. Introduction

The Variazioni Project is an eContentPlus Project funded as Content Enrichment Project with a lifespan of 30 months, starting on September 2007. The project is being coordinated by the musical private institution Fundación Albéniz and counts with several additional musical institutions (Lithuanian Academy of Music and Theatre, Koninklijk Conservatorium Brussels, Escolal Superior de Música e Artes do Espectáculo do Porto, Sibelius Academy, and Association Europeenne of Conservatoires, Academies de Musique et Musikhochschulen) and technical partners (Germinus XXI, Rigel Engineering, Exitech, Universitat Pompeu Fabra and Università degli Studi di Firenze).

The purpose of Variazioni is to provide a Content Enrichment Portal where users and musical institutions can publish, annotate and access musical contents, including its protection. In order to validate its approach, the project will provide a minimum of 700 audiovisual hours, 1000 audio hours and 2000 written documents.

Variazioni project aims at enabling the enrichment of musical content metadata provided by musical institutions and end users, and considers different types of musical contents (master classes videos, digitalised scores, etc.).

The purpose of this article is to give an overview of the Variazioni metadata model and its rationale, which have been the problems for applying traditional cataloguing systems or available standards.

2. Limitation of traditional cataloguing approaches formusic

In order to review the relevant metadata standards for Variazioni, the relevant metadata standards table for cultural heritage projects developed by the project MultiMatch [Ire07] has been updated, refined and extended for the music sector, as shown in Table 1.

After reviewing these standards [Igl08], the first conclusion is that any of the reviewed standards deal with the cataloguing of music resources with enough detail for fitting user requirements in terms of search facilities and collocations. In addition, there are important limitations in traditional cataloguing systems for music resources Traditional library cataloguing records, based on AACR2R [AacrURL] cataloguing rules and MARC [MarcURL] bibliographic and authority standards have provided a solid foundation for the required descriptive metadata elements for searching and retrieving works of music and are used

[1] This research has been co-funded by the European Community under the programme eContentPlus. The authors are solely responsible for this article and it does not represent the opinion of the European Community. The European Community is not responsible for any use that might be made of information contain within.

co-funded by the Community programme eContent*plus*

by music cataloguing agencies worldwide [Hem02]. Nevertheless, several authors have pointed out the limitations of using traditional cataloguing systems for the music domain [Mini02, Hem02].

	Schemas	Controlled Vocabularies	Projects Other
Libraries	FRBR, MARC, MODS, METS, RDA, DC, IAP	DDC, UDC, LCSH, FRAD	EDLNet OAI-PMH
Museums	CDWA, VRA, CIDOC-CRM	AAT, TGN	-
Education Sector	IEEE LOM	-	-
Audio visual sector	MPEG-7, MPEG-21	-	-
Music sector	Music Brainz	Musaurus, Music Thesarus, RILM	Variations, Music Australia, Harmos

Table 1: Relevant metadata standards for Variazioni

The main observed limitations are:
- *Lack of adequate structural* [Hem02]. Traditional cataloguing systems such as MARC lack of structural metadata which provides facilities for navigating in the internal structure of the object, such as track descriptions or time or page ranges. There are precursors of structural descriptors in AACR2/MARC, such as table of contents notes, notes about duration or the 856 tag [Hem02] for the universal resource locator (URL), but they do not allow the user to adequately search and navigate the subsections of the digitalized work.
- *Lack of adequate administrative metadata* [Hem02]. Although the MARC bibliographic record includes administrative metadata, such as copyright date, date the record was created or updated, and notes about access restrictions and file format, they are limited in scope. It is missing administrative metadata for recording technical, access rights and preservation elements.
- *Limits of the conventional on-line catalogue* [Hem02]. Search results do not group related items and users cannot take advantage of collocations. In contrast, an object oriented metadata model can improve comprehensiveness and precision of search results [Mini02], since a work can be linked to all its instantiations, roles of contributors are clearly delineated and linked to appropriate entities, etc. For example, traditional approach considers only the role author, which could be performer, composer, conductor, etc. Other examples include the title (title of the track, the container, alternative title, etc.). and the dates (date of performance, composition, record creation, etc.).
- *Impervious, pre-coordinated, multi-faceted headings* [Hem02]. The nested style of creating uniform titles and subject headings may be efficient for the cataloguer but it is often impervious to the searcher. For example, [a Sonatas. m piano. n no. 21. op. 54. r C major. o arr.] contains information about the title of the work (a), instrument (m, medium of performance), number or section (n), etc. Most catalogues do not provide separate search options for the title building blocks, and is left to the users to retrieve using keywords. Regarding subject headings, the same problem arises [Hem02]. Library of Congress music subject provides multi-faceted strings headings (such as “Sonatas (Saxophone and piano)” or “Accordion music (Jazz)”) or multi-field headings (such as “*Topical*: [Woodwind instruments. x Reeds.] *Form*: [Jazz. v Discographies]. *Geographical:* [Composers: z Austria]”).
- *Weak relationships between fields describing separate works* [Hem02]. If a record includes more than work, it is not possible to link key access points (title, performer , subject heading, etc.) to the right work, but to the whole record, restricting the search options.
- *Insufficient links between versions of a work* [Hem02]. AACR2 and MARC do provide insufficient linking facilities between versions of a work (opera, score, etc.), mainly based on uniform titles, which leads to inefficient keyword search facilities.
- *Low expressivity for musical entities*. Musical entities are described with text, which lead to introduce the same musical entity with different forms. This is the main reason to establish complex authority control rules. In contrast, a multidimensional (object oriented) model improves data accuracy and promotes its consistency, since main entities are only introduced once.
FRBR (Functional Requirements for Bibliographic Records) [IF98] has accomplished a shift in the

cataloguing area, putting emphasis on a conceptual model which is focused on the Work rather than on the Manifestation. FRBR has been applied previously in the musical domain, and new library standards, such as RDA or IAP are based on FRBR. Our conclusion is that FRBR is a good starting point for defining and modelling Variazioni metadata.. This conclusion could be considered in a wider scope. According to Gartner [Gar08], "*given the complexitiy of metadata requirements, it is perhaps not surprising that no single standard has yet emerged which addresses them all. Nonetheless, the emergence of the standards detailed in this report, all of which are based on the Functional Requirements for Bibliographical Records (FRBR) conceptual model, and the interoperability allowed by their common language, does allow for a coherent metadata landscape to be constructed on a sector-wide basis.*"
Regarding METS, METS and MPEG-21, are two standards which attempt to provide overall frameworks within which descriptive, administrative and structural metadata and have emerged from different communities [Gar08]. While METS comes from the library community (the MARC standards office), MPEG-21 comes from the multimedia community. Variazioni counts with experts in MPEG-21, and the resulting metadata will be available in MPEG-21.
The general followed approach will be based on defining the metadata model required by Variazioni partners. A **metadata crosswalk** will allow interoperability of Variazioni metadata model to be used by other communities with use a different metadata schema. In particular, for Variazioni is particularly relevant providing OAI-PMH interoperability in order to be integrated in the European Library in the future. Since OAI requires Unqualified Dublin Core metadata. A crosswalk to EDLNet metadata will be included.
Regarding the standards developed in the museum community, they deal with aspects not relevant for Variazioni (physical location or provenance of the items) and, in addition, there is an adaptation of FRBR, so-called FRBRoo, which provides an effort in modelling CIDOC CRM based on FRBR entities.

3. Adaptation of FRBR for Variazioni

This section discusses how the FRBR conceptual model can be applied in Variazioni. In order to understand better the relationship with FRBR, a first identification of FRBR entities per musical content type has been carried out as shown in Table 1.

Variazioni Contents	**FRBR 1st Group Entities**			
	W	**E**	**M**	**I**
Master class	Master Class	Master Class Event	P	MF
Score	C	Editorial Event	P	MF
Concert	C	Concert event	P	MF
Image*	Image itself (or P)	["Event"]	P	MF
Studio Recording	C	"Event Production"	P	MF
Libretto	C, 'Textual Work'	"Editorial Event"	P	MF

Table 1: Identification of FRBR entities. Legend (W)ork, (E)xpression, (M)anifestation, (I)tem, (C)omposition, MC (Musical Content) (P)roduction, MF (Media File

From this exercise, several issues have arisen:

(a) Expression and Work entities are not easy to identify in some cases, such as Master Classes or Conferences. This happens because the intellectual or artistic activity (Work) emerges while the activity (Expression) is being carried out. A similar issue has been previously reported for Western Music or Jazz improvisation in FRBRList [FRBRList] or MusicAustralia.

(b) According to FRBR, an Expression is the realization of one and only one Work entity. This can create some problems while cataloguing if the final digital file contains several Expressions (for example, a video recording with several performances or a digitalised score book with several scores, or a CD in only one track) and there is not a segmentation tool available in the system.

(c) The main Work entity in the music domain is Composition. Nevertheless, in some musical contents, such as Master Classes or Conferences, the Composition is not the intellectual / artistic activity of the Master class / Conference, but It is commonly used to exemplify a concept. They are used as subjects.

(d) Managing image and 'event material'. The image content is problematic. For example, let us consider a concert, where there are a video recording, an audio recording and photos of the event. One natural alternative is considering all of them are 'Manifestations' of the same Expression (the Concert) but recorded in different media (image, video or sound). The main problem is that the photo may not be easily linked to the performance of one particular Work, but to the general event. A similar case happens for cataloguing related material such as the announcement poster of the Concert. According to [IF06], these augmentations (illustrations, notes, glosses, etc.) of the Expression should be considered separate Expressions of their own separate works, but this makes hard the cataloguing.

(e) In digital libraries, the distinction between Manifestation and Item is not so relevant, since there is only one copy of the work (the digital media). FRBR cannot be considered as a data model, but as a conceptual schema. FRBR does not even require implementing the four entities of the first FRBR Group [IF06].

(f) While FRBR follows a top-down approach for cataloguing, cataloguing follows a bottom-up approach. Users or librarians catalogue an Item, not a Work. Users should have an easy interface in order to catalogue their media files, without being aware of the FRBR model. Expertise in implementing FRBR in standard databases [Ayr04] has shown its utility for end users to find relationships between items, which were hidden before its implementation. Nevertheless, these experiences have shown that since FRBR provides several alternatives during the cataloguing process, this can add complexity to the general understanding of the process. Some examples of these difficulties are to decide whether music and lyrics should be catalogued as different items, the definition of relationships between expressions (i.e. an interpretation (e1) based on a libretto (e2) of a work (o1)), as the cataloguing of expressions based on improvisation, such as jazz music and folk traditions.

(g) Cataloguing can be done in an iterative way. Depending on the available resources, a media file can be uploaded and catalogued with very few metadata

Based on these observations, an adaptation of FRBR for musical resources is here proposed.

Since the FRBR model has been adapted, FRBR entities has been renamed and redefined, in order to avoid confusion to the reader[2]. In particular:

- **Work** has been limited to Compositions. A **Composition** is an original piece of music.
- **Expression** has been redefined as Musical Content. A **Musical Content** (Musical Content Type) is a classification scheme of digital items which defines the nature and descriptive metadata of the digital item. Some of the musical content types identified are Master Class, Conference, Libretto, Musical Score, etc.
- **Manifestation** has been renamed as **Production**. A **Production** maintains all the metadata related to the physical edition of a Musical Content, as well as the structural metadata when the manifestation is composed of more than one Media Fragment. The structural metadata can include the order of different Media Fragments or the starting and end points of one media file with different fragments (pages, seconds, frames, etc.).
- **Item** has been renamed as Media Fragment. A **Media Fragment** is a media file or a fragment of it, and maintains all the relevant metadata of the media file, including its title and licence.

In order to clarify these elements, here follows an example of how the same items are catalogued according to standard FRBR (W: Work, E: Expression, M: Manifestation, I: Item) and Variazioni Music Application Profile (C: Composition, MC:Music Content, P: Production, MF: Music Fragment).

W1. J. S. Bach's Six suites for unaccompanied cello
- E1. Transcription for classic guitar by Stanley Yates
  - M1. Publication of the guitar transcription by Mel Bay Publisher in 1988
    - I1. Exemplar of the book in library 1.
    - I2. Separata of the guitar edition in library 1.
- E2. Performances by Janos Starker recorded in 1963 and 1965
  - M1. Recordings released on 33 1/3 rpm sound discs in 1965 by Mercury
  - M2. Recordings re-released on CD in 1991 by Mercury

In Variazioni metadata model, the structure would be as follows.

[2]A similar approach of renaming entities have been followed previously by Variations and IAP.

MC1. Score. Transcription for classic guitar by Santley Yates

- C1: J. S. Bach's Six suites for unaccompanied cello
- P1: Book edition
    - MF1: Media file of the book (page range if book includes more compositions)
- P2 Separata of the guitar edition
    - MF2: Media file of the separata

MC2. Studio Recording. Performances by Janos Starker recorded in 1963 and 1965

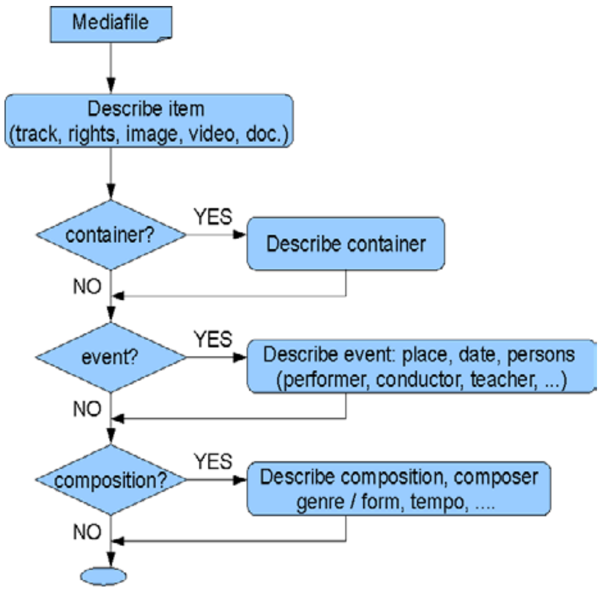
- C1: J. S. Bach's Six suites for unaccompanied cello
- P3: Recordings released on 33 1/3 rpm sound discs in 1965 by Mercury.
    - MF3: Suite 1 media file (and details of the fragment, full or time range)
        - C2: J. S. Bach Suite 1 for unaccompanied cello [is-part-of C1]
    - MF4: Suite 2 media file (and details of the fragment, full or time range)
        - C3: J. S. Bach Suite 2 for unaccompanied cello [is-part-of C1]
    - …
    - MF8: Suite 6 media file (and details of the fragment, full or time range)
        - C7: J. S. Bach Suite 1 for unaccompanied cello [is-part-of C1]
- P4: Recordings re-released on CD in 1991 by Mercury
    - MF9: Media file of the suites or details or the fragments (time range) in one media file

From this example, the main differences of the model can be outlined.

First of all, according to FRBR, and Expression has one and only one Work, and this has supposed the shift in focus from the resource (Manifestation) in the traditional cataloguing world to the Work in FRBR. Our proposal consists of modifying the cardinality of the relationship *hasWork* between Work and Expression, from 1-1 in FRBR to M-M (many-to-many). This allows solving some of the previous issues pointed out: (a) , since Compositions (Works) are not mandatory for a Musical Content (Expression); and (b), since one Musical Content (Expression) can have more than one associated Compositions (Works).

Another interesting change is the usage of the relationship *hasSubject*, in particular for linking any element of the model with *Composition*. FRBR only considers this relationship for Works. In our case, for example, for Master classes, several Compositions could be the subject (or example) of a master class. In the example previously presented, a composition can be assigned as subject of a Music Fragment, suppressing the need for a new Expression. This is depicted in Figure 1, which points out two different kind of semantic relationships between Composition and Musical Content: *isRealizedAs* and *hasSubject*. In terms of search ability, we have not found the need to distinguish between both in the implementation of the model. Furthermore, it is possible to define the subject of a media fragment, allowing a direct

Finally, the process of identifying the entities of the model is hard for end users, and a simple process for guiding the cataloguing has been defined, which is shown in Illustration 1.



llustration 1: Variazioni Cataloguing Process

8. Licence Model and Content Protection

The Variazioni project has integrated a Digital Rights Management (DRM) solution in order to control the usage of the content. In this way, Variazioni can ensure that only registered users have access to the content and thus, fulfil the content producer's requirements in this sense.

The Variazioni License Model is based on the MPEG-21 Rights Expression Language [Xing04] and considers not only the licensing from content distributors to end users, but also the step from content providers to content distributors. In other words, any content distributor that may wish to transfer or grants any right to an end user needs to own the

corresponding rights granted from the rights owner (content creator or distributor).
In the VARIAZIONI project, a content provider corresponds to the party owning the rights for a piece of work, whereas the distributor is the VARIAZIONI portal. Therefore, the VARIAZIONI portal needs to own the corresponding rights granted by the content providers in order to be able to give access rights to all its members.
Several license models have been considered during the specification of the Variazioni project:

- *PlayNoCond*. The granted user can play the content without any restriction.
- *PlayFeePerUse*. The granted user can play the content by clearing a specific fee every time the content is played.
- *PlayTimesAmountTime*. The granted user can play the content a limited number of times during a limited time interval.
- *PlayTimesInterval*. The granted user can play the content during a limited time interval.

However, since the access to enrich the content is open without any restriction to all the users registered in the Variazioni portal, the *PlayNoCond* license model has been selected for being deployed.
Table 2 depicts a license that is produced by DID:Distributor for enabling the UID:EndUser to play with no restriction the object OID:Identifier.

```
<?xml version="1.0" encoding="UTF-8"
standalone="yes"?>
<r:license
xmlns:dsig="http://www.w3.org/2000/09/xmldsig#"
xmlns:mx="urn:mpeg:mpeg21:2003:01-REL-MX-
NS" xmlns:r="urn:mpeg:mpeg21:2003:01-REL-R-
NS" xmlns:sx="urn:mpeg:mpeg21:2003:01-REL-SX-
NS"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance"
xsi:schemaLocation="urn:mpeg:mpeg21:2003:01-
REL-R-NS ../schemas/rel-r.xsd
urn:mpeg:mpeg21:2003:01-REL-SX-NS
../schemas/rel-sx.xsd urn:mpeg:mpeg21:2003:01-
REL-MX-NS ../schemas/rel-mx.xsd">
 <r:grantGroup>
   <r:grant>
     <r:keyHolder>
       <r:info>
       <dsig:KeyName>UID:EndUser</dsig:KeyName>
       </r:info>
     </r:keyHolder>
     <mx:play/>
     <mx:diReference>
       <mx:identifier>OID:Identifier</mx:identifier>
     </mx:diReference>
   </r:grant>
 </r:grantGroup>
 <!--The license is issued by the distributor.-->
 <r:issuer>
   <r:keyHolder>
     <r:info>
       <dsig:KeyName>DID:Distributor
       </dsig:KeyName>
     </r:info>
   /r:keyHolder>
 </r:issuer>
</r:license>
```

Table 2: PlayNoCond license model. The content can be used without any restriction.

The Variazioni project uses the AXMEDIS technology [AxmURL] to create protected content objects whose access is restricted to those that own a license with the corresponding access rights based on the Variazioni license model.
For this purpose, the Variazioni portal is linked to the AXMEDIS DRM servers so that whenever a protected object is generated, the corresponding licenses are automatically produced to grant all the Variazioni registered users the access right.
The protected objects can be used by any user:

- that is registered on the AXMEDIS servers, i.e. that owns a personal user certificate;
- that owns an AXMEDIS player, which is installed and certified, i.e. which has been linked to the user and device by means of an automatic and transparent process given the user certificate.

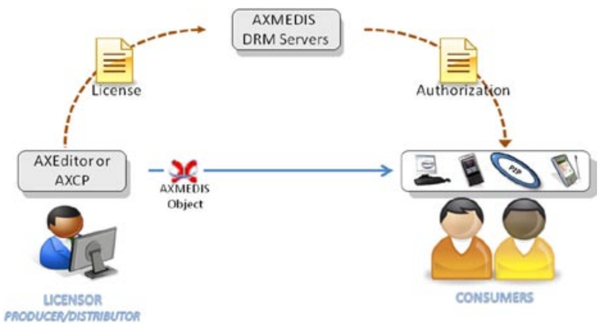


Illustration 2: Usage of the AXMEDIS technology in Variazioni for content packaging, protection, distribution and consumption.

The protected content objects can be then accessed by users by means of any of the AXMEDIS players for PC, PDA, STB, mobile, etc. The AXMEDIS ActiveX Player can be used to integrate the AXMEDIS player into any HTML page, thus making the integration simpler. The AXMEDIS players can be downloaded for free from the AXMEDIS Portal [AxmURL].

9. Conclusions and Future Work

The web2.0 user participation along with the new technological advances define a new landscape where metadata plays an important role for content search ability and exploitation. Musical assets have been inadequately catalogued with traditional standards, and there is a need for defining more precise metadata schemas for musical resources.

This article presents a novel model, based on FRBR, for musical resources which has been formalised as a Dublin Core Application profile, and has been implemented in the Variazioni project [VarURL].The main advantages of the model are its ability for collocated contents and navigation within the metadata model.

In addition, a flexible licence model has been formalised in MPEG-21 Rights Expression Language and implemented with Axmedis Platform.

Our ongoing work is the validation of the model with end users, since this model has been validated with musical analysts from the musical institutions which participate in Variazioni.

10. Acknowledgments

The authors wish to thank all the partners of Variazioni for their effort and contributions along the project. We particularly wish to thank our colleagues Daniel Molina, Saúl Navarro, David Jiménez, Santiago González, Mari Cruz Mansilla, Rui Quintas and Piero Alcamo for their enthusiasm and proactiveness in the implementation of the Variazioni Metadata Model. The authors also wish to thank Pablo Clemente for his detailed review of this article,

11. References

[Ayr04] Case Studies in Implementing Functional Requirements for Bibliographic Records [FRBR]: Auslist and MusicAustralia, Marie Louise Ayres, Austalian Library Journal, number 1, vol. 54, 2004. Available at http://www.alia.org.au/publishing/alj/54.1/full.text/ayres.html.

[AxmURL] Axmedis Project Official Web Site, available at http://www.axmedis.org.

[AacrURL] Anglo-American Cataloguing Rules (AACR) official web site, available at http://www.aacr2.org/.

[Ca06] Casey, Michael and Savastinuk, Laura. Library 2.0. Service for the next-generation library. Library Journal, January 2006

[IF06] FRBR Chapter 3: Entities, Proposed changes to the FRBR Text by the IFLA Working Group on the Expression Entity, August 2006.

[IF98] International Federation of Library Associations and Institutions, IFLA, Functional Requirements for Bibliographic Records, 1998. Available at: http://www.ifla.org/VII/s13/frbr/frbr.pdf

[FRBRList] FRBR Mailing list, moderated by Patrick le Boeuf. List address FRBR@bnf.fr

[Gar08] Metadata for digital libraries: state of the art and future directions, Richard Gartner, JISC Technology & Standards Watch, 2008.

[Hem02] Why not MARC? Harriette Hemmassi, Proceedings of the 3rd International Conference on Music Information Retrieval, ISMIR 2002, p. 242-248, 2002.

[Igl08] D2.3. Updated Variazioni Musical Metadata Definition, Variazioni project, 2008.

[Ire07] Capturing e-culture: Metadata in Multimatch, Neil Ireson and Johan Oomen, in Ontology-Driven Interoperability for Cultural Heritage Objects, Working Notes, DELOS-MultiMatch Workshop, Tirrenia, Italy, February 2007.

[MarcURL] MAchine Readable Cataloguing, developed by the Library of Congress, official web site available at http://www.loc.gov/marc/.

[Mini02]A Digital Library Data Model for Music, Natalia Minibbayeva and John W. Dunn, in Proceedings of the Second ACM/IEEE Joint Conference on Digital Libraries, pp. 154-155, 2002.

[VarURL] Variazioni project available at http://www.variazioniproject.org

[Xing04] MPEG-21 Rights Expression Language: enabling interoperable digital rights management, Multimedia IEEE, vol 11, issue 4, Oct-Dec, p. 84-87, 2004.

SEMANTICS AND ONTOLOGIES FOR MULTIMEDIA OBJECTS REPRESENTATION AND METADATA MANAGEMENT IN SOUND ARCHIVES

Francois Scharffe[1], Michael Luger[1], Yves Raimond[2], Ivan Damnjanovic[2], and Josh Reiss[2]

[1]STI Innsbruck, Institut fur Informatik
21a Techniker Strasse
6020 Innsbruck Austria
{firstname}.{lastname}@sti2.at

[2]Queen Mary, University of London
Mile End Road, E1 4NS
London, United Kingdom
{firstname}.{lastname}@elec.qmul.ac.uk

ABSTRACT

Sound archives have been massively digitalized in the past twenty years. We are also witnessing that many of them are becoming available on-line. The emergence of the web, and its evolution towards the semantic web opens a new phase for the publication of digital archives. The data and assets they contain can be made available in a structured way, providing more precise, as well as wider querying possibilities. In this paper, we present an ontology for easily publishing and managing digital archives, based on semantic web technologies. An architecture based on the Music Ontology is successfully being used within the EASAIER (Enabling Access to Sound Archives through Integration, Enrichment and Retrieval) European project.

Index Terms— Sound Archives, Multimedia Retrieval, Music Ontology.

1. INTRODUCTION

Ontologies are the backbone of the semantic web in particular, and of modern knowledge representations in general. An ontology provides a way to describe a restricted world we are in a logical language (description logics, and in the semantic web context, OWL (which can be serialized as XML)), allowing automatic reasoning. It is far more than just a metadata scheme (descriptors attached to top-level nodes), as the raw MM file is just an object which has the same relevance as any other objects (such as a particular artist, a particular performance, and so on...). An ontology answers the following use-cases:

- **Automatic reasoning** - An ontology, by being formally specified, allows automatic reasoning on objects in the described domain. For example, It is possible to query an ontology-based system for all recordings involving wind instruments and gain access to those involving flute, oboe and not only the ones directly “tagged” with wind instrument.
- **Cross-Media knowledge management** - Each multimedia object is relevant, and described in a semantic graph. By using an ontology, a user can access both the video of a performance and the related recording, as well as the lyrics.
- **Flexible knowledge representation** - For example, using an ontology, you can perfectly recognise the existence of an object representing a particular performance of a piece, without the related recording. This is impossible with a standard metadata approach.
- **Distributed multimedia repositories** - Using OWL, multimedia files are identified by an URI. It means that files can be on a FTP server, on an HTTP one, accessible through SSH, streamed, or even on a peer-to-peer network. The corresponding URI just has to be resolvable.
- **Exporting multiple metadata standards / MPEG7 link** - By building a particular interpretation of the theory held by the ontology, it is possible to export some knowledge in several metadata standard. From really poorly expressive ones (ID3,...) to highly expressive ones (MPEG7).

2. MUSIC ONTOLOGY

The Music Ontology [1] is built on top of the Timeline ontology [2] and the Event ontology [3], as well as the Functional Requirements for Bibliographic Records ontology (FRBR) [4], mainly used for its concept of **Work** (an abstract, distinct, artistic creation), **Manifestation** (physical embodiment, like a record, for example), and **Item** (a single exemplar of such a manifestation, like a particular vinyl). We also use the Friendof-a-friend ontology (FOAF) [5], and its concepts of **Person** and **Group**. We define a number of music-specific concepts, on top of these three ontologies.

On top of FRBR, we define **MusicalWork**—an abstract musical creation (such as Franz Schubert’s Trout quintet), **MusicalManifestation**, which can be a a **Record** or a **Track** among others), and **MusicalItem**, which can be a **Stream**, a particular **CD** or a particular **vinyl**, etc. On top of the FOAF ontology, we define **MusicArtist** and **MusicGroup**.

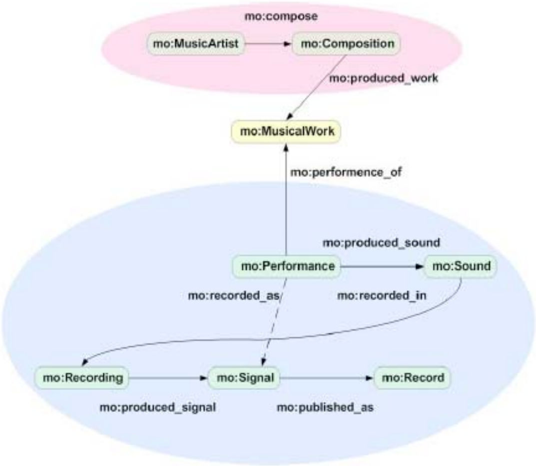


Figure 1. Music Ontology Workflow

On top of the **Event** ontology, we also define a number of concepts, relative to the music creation work flow. **Composition** deals with the creation of a **MusicalWork**. **Arrangement** deals with an arrangement of a **MusicalWork** and can have as a factor a **MusicalWork**, as an agent an **Arranger** and as a product a **Score**. **Performance** denotes a particular Performance, and can have as factors a **MusicalWork** and a **Score**, a number of musical instruments, equipments, and as agents a number of musicians, sound engineers, conductors, listeners, etc. A **Performance** can have as a product another event: **Sound** — a physical sound. This sound may itself be a factor of a **Recording**, which may produce a **Signal**. This **Signal** can then be published as a **MusicalManifestation**. This leads to a work flow depicted in Figure 1.

The feature ontology [6] aims at creating a generic framework for expressing features of audio signals (Mel Frequency Cepstral Coefficients, chromagram, onsets, etc.). It uses the broad definition of the Event concept in order to express an artificial classification of a time region, corresponding to a particular feature. Therefore, it defines a sub class of Event: FeatureEvent, allowing to classify time regions corresponding to features.

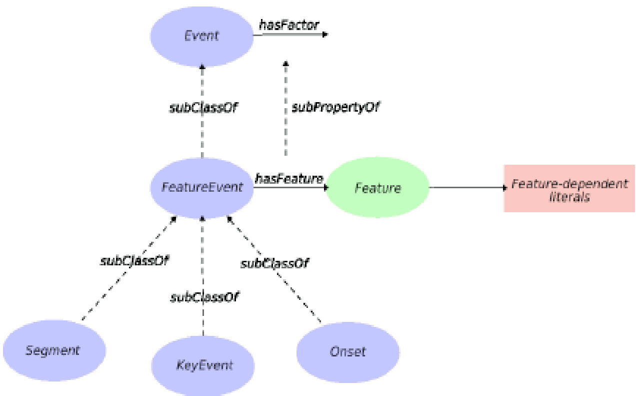


Figure 2. Features Ontology

Feature Event may have a number of Feature factors, representing a particular feature, such as a chromagram or a key (Figure 2).

Linking Open Data on the Semantic Web - As an example of such a linking, we may provide information about a festival happening in Montreal, Canada on 28 June 2007. We can link our Festival instance using the event:place property to its geographical location resource in Geonames. A user agent crawling the web of data can then jump from our knowledge base to the Geonames one, by following this link, and get detailed information about the place where the festival is happening.

3. CONCLUSIONS

The music ontology has a quickly growing users community, and can be considered as the reference ontology for publishing audio archives on the semantic Web. An architecture based around the music ontology that can be reused to integrate sound archives, and with extending the ontologies, to integrate any media archive, and publish its contents on the semantic Web. This gives a powerful tool for archivists willing to exploit the rich knowledge contained in archives, and give access of this knowledge to a wider audience [7],[8].

ACKNOWLEDGMENTS

This work has been partially supported by the European Community 6th Framework Programme project EASAIER (IST-033902): www.easaier.org.

REFERENCES

[1] Y. Raimond, S. Abdallah, M. Sandler and F. Giasson, “The Music Ontology”, Proc. of the International Conference on Music Information Retrieval, 2007

[2] Y. Raimond and S. A. Abdallah, “The Timeline Ontology”, 2006

[3] Y. Raimond and S. A. Abdallah, “The Event Ontology”, 2006

[4] D. Brickley and L. Miller, “FOAF Vocubulary Specification”, 2005

[5] I. Davis and R. Newman, “Expression of Core FRBR Concepts in RDF”, 2005

[6] Y. Raimond, “The Musical Features Ontology”, 2007

[7] F. Scharffe, Y. Raimond, L. Barthélémy, Y. Ding, and M. Luger, ”Publishing and Accessing Digital Archives using the EASAIER Framework,” Workshop on cultural heritage and the semantic web. ISWC, Busan, Korea, 2007.

[8] F. Scharffe, M. Luger, Y. Raimond, “Semantics in the EASAIER Framework,” Enabling Access to Sound Archives, workshop at Minerva conference, Jerusalem, 2007.

What is MILE? (http://www.mileproject.eu/home)

Andrea de Polo
Fratelli Alinari Firenze
andrea@alinari.it

Abstract

MILE (Metadata Image Library Exploitation) *wants to make art available to everyone by improving metadata. MILE is an EC funded project which aims to improve the use, accessibility and trade of digital images throughout Europe.*

Image libraries maintain a vast wealth of Europe's cultural heritage in their archives of reprographic transparencies of original works of art. In the rush to digitise this content to keep up with increasing technological demands, the systems used to create the metadata supporting these images have struggled to sustain effective functionality and accessibility, thereby restraining its maximum potential for exploitation through the EU.

MILE was chosen for funding by the EC from over 300 project proposals under the eContentplus call of the i2010 digital libraries initiative, to preserve and promote European cultural heritage.

1. The areas of the project

The project is divided into three core areas of investigation:

1.1. Metadata Classification Why focus on metadata classification?

In visually-based cultures where supply and demand for digital images is growing rapidly, the need for swift, efficient and cost-effective image cataloguing systems is ever increasing. Image collections, art librarians and museum curators have adopted various schemas for cataloguing visual art images, none of which have been designed specifically for cultural collections.

1.2. What will MILE do?

MILE brings together partners who include metadata creators, technology providers and end users to establish the challenges in creating a harmonised cataloguing system by studying the problems with existing standards. This will help MILE to find solutions for better cataloguing systems. In doing so, MILE will consider guide models such as the 'Cataloguing Cultural Objects' project and MDA's SPECTRUM standard, as a means of providing a potential solution for implementing not only harmonisation across Europe's image archives, but also the maintenance of high quality and authoritative standards.

1.3. Metadata Search and Retrieval What is metadata search and retrieval?

Many users of digital image libraries suffer from difficulties of access through unsuccessful search mechanisms or language barriers. For example, if the name of an artist is entered incorrectly – if the artist is not well known to the end user, or if the user is not searching in his mother-tongue language - the search yields no results. Another major problem is how to translate metadata into another language whilst maintaining effective search and retrieval results. For example, when one image library translated a set of images from English into German, the metadata for 'oil on canvas' read in German 'benzin auf leinwand', where the literal meaning of 'benzin' is petrol - literally, lost in translation.

1.4. Translation?

And although the majority online language is English, other languages such as Chinese, Spanish and Japanese represent a significant proportion that cannot be ignored if we are to fully exploit European Cultural heritage in the form of digital images and improve the generation of income.

1.5. Thesauri?

Various projects working with thesauri to provide alternative meaning lists to allow for the multiple meanings of some words in different languages - such as the MICHAEL Project - have made significant headway in improving translation of metadata. However, we are still some way off from a reliable translation scheme or thesaurus specifically created for images.

1.6. What will MILE do?

MILE will survey, analyse and collate information about the factors inhibiting the production of reliable translation options and thesauri. Academic and technical experts from all over the EU, including Alinari, System Simulation, CityPassenger, Archetypon and Trinity College Dublin will take part in seminars and discussion groups, to evaluate translation systems and thesauri, and research results from ongoing efforts such as the MICHAEL Project. This will enable MILE to recommend strategies for more comprehensive and reliable multilingual access to digital images.

2. Intellectual Property Rights as Metadata

2.1. What does metadata have to do with IPR?

The metadata attached to a digital image may include information about the artist, date of creation, copyright holder and/ or license holder. Many images cannot be made publicly available on the Internet without the presence of this information (metadata). If this metadata does not appear with the image, the image user may be liable for copyright infringement.

2.2. What is the problem?

" Copyright or Intellectual Property law is not standardised across Europe so image holders and image users face a variety of rules depending on where they are situated.

" There are often many layers of copyright within an image, all of whose rights need to be considered. This may include the photographer, the creator of the original work, the location which holds the original work and/or other rights holders. Determining all the rights holders within an image requires an understanding of the source of the image, the content portrayed and the creation of the image. Much of the time, all of the information is difficult to locate.

" Image libraries are torn between their duties to protect the rights holders of the images they hold whilst trying to promote these images as widely as possible to maximise the sale of these images so that the rights holders may earn an income from their images.

" In some cases, copyright information about the rights holder(s) of an image is not known. These images are referred to as 'orphan works'. The number of these works in the EU has not yet been quantified but The Bridgeman Art Library alone receives requests for at least 300 'orphan works' per year so the potential income loss is substantial.

" Protecting digital images from illegal reuse is another crucial consideration for any online image archive. Piracy protection may include technological solutions which are often expensive but these solutions are often varied and fragmented and many offenders are not even aware that they are doing anything wrong.

The MILE project is working towards 4 key objectives regarding IPR:

1. Investigate and document the IPR problems and issues facing digital image users and providers throughout the European Union and raise awareness of these problems / issues.
2. Discuss and evaluate potential solutions to the above problems within our project network of skilled experts.
3. Provide best practice recommendations both to image users and image providers, and offer recommendations on future IPR legislation to the EC Parliament in the hope of achieving a degree of harmonization throughout the EU.
4. Produce an 'Orphan Works' database which will enable image holders to post their own 'orphan works' and the public to post information related to the images in the hope of attaining copyright information about these works, as well as acting as a due diligence exercise case study.

3. How will MILE help?

As a result of the discussions and seminars which will focus on the 4 objectives above, MILE will produce a guide to licensing processes, best practices and standards in IPR metadata for digital images. MILE will work with partner trade associations to promote and disseminate this guide as widely as possible. This will serve to raise awareness and elucidate IPR procedures for all European citizens so that they may use and exploit European culture through digital images, more readily, more safely and with more understanding.

Credit:
MILE consortium

EDCine Project for Digital Cinema and Archives - Digital Preservation and Access of Film and Digital Cinema Material and new developments in audio reproduction

Organized by

Fraunhofer Institute for Integrated Circuits, Germany

Coordinated by

Arne Nowak, Fraunhofer IIS (Germany)

AXMEDIS 2008: proceedings of the 4th International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution. Workshops, Panels and Industrial Applications. Florence, Italy 17-19 November 2008
edited by Paolo Nesi, Kia Ng, Jaime Delgado
ISBN: 978-88-8453-811-6 (online). ISBN 978-88-8453-810-9 (print). © Firenze University Press, 2008

Preservation and Access solutions for Moving Image Archives in the Digital Cinema era

Nicola Mazzanti
Consultant, Cinémathèque Royale de Belgique
Arne Nowak
Fraunhofer Institute for Integrated Circuits IIS
Paul Read
Consultant, Paul Read Associates

Abstract- **Within the context of the EU-funded project EDCine, a solution for a system allowing digital preservation and multi-quality, multi-format distribution and access to high quality moving image content is proposed, based on an architecture consistent with OAIS reference model, and standardized file formats.**

I. Introduction

Moving image archives play a decisive role in preserving the cultural heritage of our society and in making it accessible for cultural, educational as well as commercial purposes. Publicly funded archives include national and regional institutions with a remit to collect, preserve and provide access to collections relevant to their country, region, area or special interest. Commercial archives are more diverse and include stock-shot licensing companies, newsreel and feature film collections, producers, distributors as well as rights holders. Many film studio collections are in the process of diversifying into digital post-production and digitally borne input, as well as content destined for digitally projected cinema, home video and home cinema, internet and television distribution. Public archives are growingly concerned with the preservation of digitally borne (or post-produced) works, and are exploring new modes of access to their collections, including various digital channels (digitally projected cinema, home video and home cinema, internet and television distribution).

Film archives may store their analogue film collections under strictly controlled low temperature and humidity conditions that can provide a life expectancy of several hundred years for film already in the early stages of decay, and thousands of years for fresh film (this according to the most recent research, mostly by the Image Permanence Institute [1]). However, in the current climate of increasing digital presentation, film display except in cinemas (and this too may become untenable in time) is increasingly considered insufficient to respond to users' demand for broader access to collections.

Moving image collections of all sorts require easy and economical access to digital versions of many different qualities and file sizes both from analog (film, analog video) in their current collections, and, increasingly, from digital content. Digitally borne content, whether new or derived from film, will also require practical, secure and affordable digital preservation procedure.

The three partners (Cinémathèque Royale de Belgique, Fraunhofer IIS-Institute for Integrated Circuits, and MOG Solutions) of the archives-related section of EDCine project (funded by the 6^{th} Framework Program for Research, Technological Development and Demonstration of the European Commission) intended to meet all these requirements by presenting a two-tier storage model that provides a framework for both digital preservation at best quality, and uncomplicated access to the stored items [2].

From a system architecture's standpoint, the EDCine Archive System proposes a solution based on the OAIS reference model (a conceptual framework for digital repositories for the preservation and delivery of digital or digitised content, an ISO standard [3]), and on the definition of two different file formats to store image, audio and metadata. Each of these file formats is able to store archived items in the best possible quality (at this time), and to facilitate the access to the archived items into many different distribution formats. A modular implementation assures scalability and provides interfaces to existing systems and the future addition of new functions.

The proposed architecture consists of two packages, the Master Archive Package (MAP) for long-term preservation, and an Intermediate Access Package (IAP) designed to make the access to the stored items faster and simpler. Following OAIS reference model, both MAP and IAP are designed as information packages where the content (image, sound, texts, etc.) is stored jointly to its technical metadata to ensure that the content is correctly displayed when accessed. This is obtained by using MXF (Material eXchange Format) as a wrapping format.

II. Archival requirements

Digitisation for preservation and access being a costly and labour-intensive, archives acknowledge that viable solutions must allow for high-quality preservation so that digital content can be easily distributed on multiple platforms, ranging from very low (streaming) quality to high end distribution (broadcast, D-Cinema) in other words, all moving image film, video and digital archives need solutions that can store all or most of their images (and related sound) while providing access to them in an increasing variety of formats and qualities, with a multi-channel, multi-platform approach.

Although EDCine's principle focus is on Digital Cinema and on cinema quality images, its overall concept and architecture are designed to serve a wide variety of moving image and related sound content, including analog and digital video, and film images at lower quality than film.

Clearly, in a context of fast-changing technological environment, no new concept or process can ever be "future proof", but the EDCine Archive System's implementation offers a solution that meet current users needs while being able to fit (or to adapted to fit) into a future environment with new requirements.

As the expected decline in film projection in the cinema occurs, archives, cinematheques, specialist art house cinemas, and any cinema planning to show archival films will increasingly need to display a digital version, as well as to distribute quality on multiple platforms. In most instances, digital projection will require to be compliant with an eventual world standard (currently under consideration by SMPTE). Other digital cinema display methods are already a reality across the world and may need to be taken into account in time.

Film for cinema has always generated its own unique experience in the audience, which has varied with time, location and technology for over 100 years. Archives, cinematheques and specialist cinemas require that the cinema projection of heritage films (best defined as films shot and released prior to digital projection becoming used or standardized) be authentic and as faithful a representation on the screen of the projected original film as possible (a requirement already observed in the restoration of old films).

The following list (a summary of the details provided by the Cinémathèque Royale in consultation with the Fiaf Technical Commission, to the research partners) records some of the characters that provide this authenticity, but is not exhaustive:

1 The resolution of the projected screen image should not be visually lower than that of the original film image (it is appreciated that this requirement is difficult to quantify).
2 The frame rate of a digital cinema projection should be the same as that of the original film.
3 The aspect ratio of the image should be that of the original film
4 If appropriate to the original period, a film programme of mixed aspect ratio content should be shown using common height principles.

Basic requirements of a digital storage system for film, in brief and as proposed to the research partners include:

1 Ingest should be from both film images, film sound, and any analogue or digital video or data version of a programme, images or sound.
2 For access purposes, the Intermediate Access Package (IAP) must be capable of producing the wide range of current formats and media output, defined as Distribution Access Packages (DAP) in the EDCine Archive System.
3 The MAP, the long term digital storage format, which will in time be economically and practically viable, must store data in a lossless format.
4 D-Cinema output versions, as any other high quality output versions of the future that are generated from long term digital storage formats should be as close to the original film, or digital version, as possible. Hence conversions of colour space, resolution and frame rate, etc., are to be avoided, or if unavoidable should be losslessly and accurately reversible.
5 Film images will not always be in a restored form when ingested for long term storage as the MAP, and therefore if restoration is deferred, extensive image and sound manipulation must still be possible subsequently. This has several implications, the main concerning resolution and bit depth of archived content.
6 A single "ingest standard" would also be widely welcomed by all archives facing the problem of managing and conserving content deposited in a wide range of different formats, as for example archives that manage a legal deposit obligation.

III. Proposed System Architecture and File Formats

An archival system architecture for preservation and multi-format access of high quality moving image content must provide a solution for two major use cases with partly oppositional requirements.

On the one hand these are the long-term archival, with the requirement to store the source images and sound without any loss of information. This practice usually requires lossless compression and results in large amounts of data, due to the high spatial resolutions used in cinematic production.

On the other hand there is the requirement of frequent access to archived items. This practice usually does not require the original lossless quality of the source images and data. Instead the focus lies on easy and standardized access through local and remote channels. This results in the requirement for a significant (and often extreme) reduction of the original amount of data and the restriction of certain coding parameters to ensure maximum compatibility with a wide range of decoding and playback equipment. To illustrate this point, Table 1 illustrates the requirements in terms of data rate, and

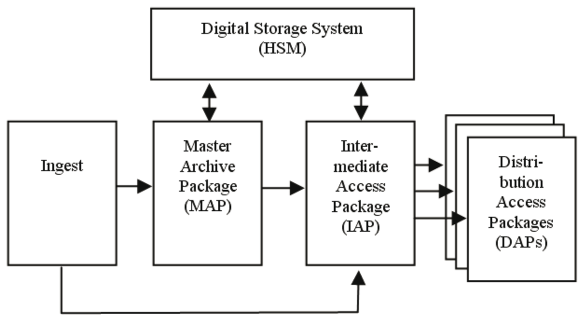


Fig. 1: Overview of the EDCine Archive System Architecture

consequent compression rates, for some of the formats managed by the EDCine Archive System, from the input of the digital master of a digitally produced film (Digital Intermediate) to an HDTV distribution format.

TABLE 1
DATA RATES IN PRODUCTION AND ARCHIVE WORKFLOW

Type	Data rate	Compression Rate
Digital Intermediate 4096x2160@24fps uncompressed 16Bit, 3 comp.	1.2 GByte/s	1:1 No Compression
MAP 4096x2160@24fps 2:1 compressed	600MByte/s	2:1 Compression
IAP (4096x2160) @24fps 500MBit/s data rate	500Mbit/s	12 Bit per Comp. + 19.2:1 Compression
HD DVB-S2 (1920x1080) @24fps 10MBit/s data rate	10MBit/s	Downscaling + 420 Subsampling + 8 Bit per Comp + 80:1 Compression

The above requirements lead to a two-tier system architecture with different file formats and JPEG 2000 profile specifications for long-term storage and for frequent access (for distribution, browsing or display).

Archived material may be stored in either one of the two formats or in both formats simultaneously depending on the type and usage scenario for the material. The long-term storage format should be used if cinematic material needs to be preserved digitally in all of its aspects. The access format shall be used if access to the material is the main concern. The latter can be generated through transcoding from the former.

The access format is close to the formats standardized by SMPTE DC28, for simplified distribution of digital content to cinemas. It can also be used to transcode to other formats for delivery to end users. For more information, see section 3.

Cinematographic content usually consists not only of image sequences but also additional data of various types. Typically, these are one or more sets of single or multi-channel audio data, timed text for subtitles and technical (optionally, also descriptive) metadata of different details levels. All of this data should be stored together in one place as described by the asset store approach standardized within OAIS.

In order to meet all the above requirements, as well as those of using only standardized and non-proprietary formats (a key requirement to ensure long term conservation of data), JPEG 2000 was chosen for both lossless and lossy compression, and MXF was selected as a wrapper to package together all data streams and metadata in the MAP and IAP.

The key specifications of the Master Archive Package (MAP) are within the limitations of JPEG2000 and AES recommended practices:

- Image resolution up to 16K (16384 x 8640)
- Any image frame aspect ratio
- Any image bit depth (only limited by JPEG 2000 maximum bit depth, which is for practical implementations 32 bits per component)
- Any colour space
- Any frame rate
- Image components up to 8
- Mathematically lossless compression for image content (can also be lossy if required, e.g. when archiving an already compressed DCP)
- Audio data in RAW format (optional MPEG-4 SLS)
- No audio sampling restrictions (at least support of recommended values as in AES-5)
- No restrictions for word length in audio (at least 16 bit or 24 bit)
- Discrete (i.e. no matrix encoding) audio channels (unrestricted number)
- MXF wrapper. No limitations on the Operational Patterns (implementation is currently restricted to Op1a and Op1b).

The key specifications of the Intermediate Access Package (IAP) are within the limitations of JPEG2000 and AES recommended practices:

- Image resolution up to 2k (2048 x 1080) or 4k (4096 x 2160) (depending on adopted profile)
- Any image frame aspect ratio
- Image bit depth up to 12 bits
- Any frame rate
- Image components: up to 3
- Compression aimed to produce a maximum bit rate to match the highest requirements for distribution (at the moment those for D-Cinema, whose standards are currently being finalized)
- Audio data in RAW format (optional MPEG-4 SLS)
- 48kHz or 96kHz audio sampling frequencies
- 16 bit or 24 bit word length for audio
- Discrete (i.e. no matrix encoding) audio channels (up to 16 channels)
- MXF wrapper. No limitations on the Operational Patterns, (implementation is currently restricted to Op1a and Op1b).

JPEG-2000 is an open standard and makes possible both mathematically lossless, and lossy (but visually lossless) compression. An added advantage of this codec is the compatibility with the intended SMPTE digital cinema projection standard, which is based on JPEG2000. The JPEG2000 profiles for digital cinema (Profile 3 and Profile 4) are already an ISO standard (ISO/IEC 15444-1:2004/Amd 1:2006), and they are referenced by SMPTE 429-4-2006.

Audio data shall be stored in a lossless manner in both MAP and IAP. For this purpose RAW audio format shall be used. As an option audio data can also be represented as MPEG-4 Scalable Lossless Coding (SLS) specified in ISO/IEC 14496-3:2005/Amd 3:2006 as a compressed but lossless audio format. All audio channels shall be organized as discrete channels. All channels of a multi-channel representation shall have the same properties. In the MAP, the number of channels is unrestricted, while it is limited to up to 16 in the IAP. The mapping of the audio channels is described in the metadata section and is independent of the order of the channels.

MXF is a wrapper file format for the exchange of audiovisual material and related metadata, and in the EDCine project is utilized to store the JPEG 2000 compressed image sequences together with any accompanying data (audio, text, etc.) and metadata providing the synchronization between the separate elements of the essence: images, sound tracks, data, etc. Metadata is to be stored together with image and other media data in the same file, although it is possible to mirror metadata in an external database to simplify search and retrieve functions, and in this case synchronisation of metadata will be guaranteed by the system. Metadata can consist of any combination of structural, descriptive and historical metadata. Structural metadata shall be stored in a format appropriate for the chosen file format. In all cases there shall be a minimum set of structural metadata stored. For example, for image data this shall consist of all information relevant to the employed JPEG 2000 profile and contain at least image size, frame rate, colour space, sub-sampling information, number and meaning of components and bit depth. Descriptive and historical metadata shall be stored as human readable text in a UTF-8 XML representation or using the metadata storage mechanisms provided by the file formats.

IV. System implementation

In order to ensure a modular and upgradeable system, the EDCine Archive System adopts a SOA (Service-Oriented Architecture), which allows a flexible number of processing and encoding/decoding modules. A SOA client allocates tasks to SOA services/modules, which can then process the tasks on single or multiple computers. Thus, a scalability and extensibility of processing modules and computing processors is made possible: new source and output formats can be handled in the future by adding new services; tasks can take advantage of enhanced performances provided by hardware accelerators, GPUs or multicore processors, whenever these are available.

Currently the EDCine Archive System is in its implementation phase, with several key modules already completed (as transcoding modules for J2k to H264, generation of MAP and IAP formats, conversion from MAP to IAP and vice versa), and others still in progress. These modules will be integrated in a first demonstrator (scheduled for the second quarter of 2009) designed to illustrate the functionalities of the system as well as some of the many workflows the system can support.

As we mentioned earlier, the EDCine Archive System' modular architecture was designed to accommodate a wide range of workflows to meet the needs of archiving and distributing both archival content and current, digital productions. The workflows identified in EDCine's initial phase (with the input from archives and digital post-houses) indicates the need for the system to ingest, manage and preserve a wide range of input formats (digitised analogue film and video, digital video and Digital Cinema products) and to transcode them on demand for distribution and access via different digital formats. As an example of the many possible workflows, Figure 2 shows how the system handles the ingest of digitally post-produced film (a Digital Intermediate), its conversion in an MAP and subsequently in an IAP, and from this the production of three delivery formats, an Internet-browsing format, a Digital Cinema Package and an HDTV master. As the figure shows, the System will handle the necessary compressions, transcoding, color conversions, as well as management of the relevant metadata.

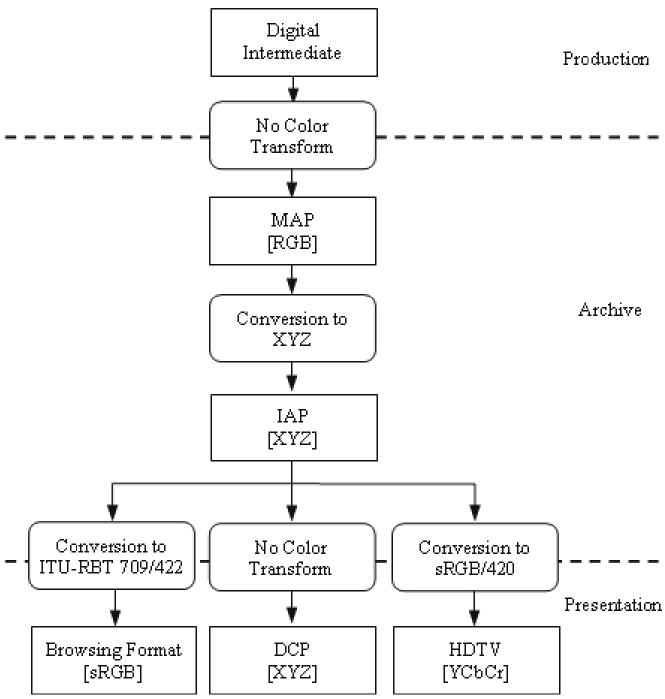


Fig. 2. Workflow for ingest, archiving and distribution of a digitally post-produced film

V. Standards proposals

It is part of EDCine/Archives' mandate to undertake all the necessary actions to ensure that all the relevant parts of the EDCine Archive System either utilize existing standards or whenever this is applicable, standardization processes are undertaken.

More specifically, the standardization process for three new archive-related profiles for JPEG 2000 was started, and actions for MXF are currently being assessed.

Table 2 contains a description of a proposal for new standardized JPEG2000 profiles, as it stands in March 2008. Currently the proposal is being discussed and reviewed, so details can obviously change.

References

[1] Bigourdan, J.-L., “ Film Storage Studies—Recent Findings,” *Preserve—Then Show, 60th Anniversary Seminar of the Danish Film Institute/Film Archive*, Copenhagen, November 11-13, 2001. Other references can be found on the I PI website *(http://www.imagepermanenceinstitute.org/)*.

[2] EDCine/Archives web page at http://www.ledoux.be/fr/edcine.htm

[3] ISO Standard nr. ISO 14721:2003. Reference Model for an Open Archival Information System (OAIS), CCSDS 650.0-B-1, BLUE BOOK, January 2002, Consultative Committee for Space Data Systems, http://public.ccsds.org/publications/archive/650x0b1.pdf

TABLE 2
CODESTREAM RESTRICTIONS FOR ARCHIVE APPLICATIONS

	Scalable 2k digital cinema profile (Intermediate Access Package)	Scalable 4k digital cinema profile (Intermediate Access Package)	Long-term storage profile for cinematic content (Master Archive Package)
SIZ marker segment			
Profile Indication	Rsiz=5	Rsiz=6	Rsiz=7
Image size	Xsiz <= 2048, Ysiz <= 1080	Xsiz <= 4096, Ysiz <= 2160	Xsiz <= 16384, Ysiz <= 8640
Tiles	one tile for the whole image: YTsiz + YTOsiz >= Ysiz XTsiz + XTOsiz >= Xsiz	one tile for the whole image: YTsiz + YTOsiz >= Ysiz XTsiz + XTOsiz >= Xsiz	One tile for the whole image or minimum tile size: YTsiz + YTOsiz >= 1024 XTsiz + XTOsiz >= 1024
Image and tile origin	XOsiz = YOsiz = XTOsiz = YTOsiz = 0	XOsiz = YOsiz = XTOsiz = YTOsiz = 0	XOsiz = YOsiz = XTOsiz = YTOsiz = 0
Sub-sampling	$XRsiz^i = YRsiz^i = 1$	$XRsiz^i = YRsiz^i = 1$	No restriction
Number of components	Csiz = 3	Csiz = 3	Csiz<=8
Bitdepth	$Ssiz^i = 11$ (i.e., 12 bit unsigned)	$Ssiz^i = 11$ (i.e., 12 bit unsigned)	No restriction
RGN marker segment	Disallowed, i.e., no region of interest	Disallowed, i.e., no region of interest	Disallowed, i.e., no region of interest
COD/COC marker segments	Main header only	Main header only	Main header only
Coding style	Scod, Scoc = 0000 0esp, where e=s=0, and p=1 Note – e=0: EPH marker shall not be used s=0: SOP marker shall not be used p=1: precincts defined in $SPcodI^i$ / $SPcocI^i$	Scod, Scoc = 0000 0esp, where e=s=0, and p=1 Note – e=0: EPH marker shall not be used s=0: SOP marker shall not be used p=1: precincts defined in $SPcodI^i$ / $SPcocI^i$	Scod, Scoc = 0000 0esp, where e=s=1, and p=0 or 1 Note – e: EPH marker shall be used s: SOP marker shall be used p: precincts with PPx=15 and PPy=15 or defined in $SPcodI^i$ / $SPcocI^i$
Progression order	CPRL	CPRL	CPRL
Number of layers	L=2	L=2	L<=5
Multiple component transform	No restriction	No restriction	No restriction
Number of decomposition levels	$N_L <= 5$ Every component of every image of a distribution shall have the same number of wavelet transform levels.	$1 <= N_L <= 6$ Every component of every image of a distribution shall have the same number of wavelet transform levels.	No restriction, with respect to: (Xsiz-XOsiz)/D(I) <= 64 (Ysiz-YOsiz)/D(I) <= 64 and D(I)=pow(2,N_L) for each component I
Code-block size	xcb = ycb = 5	xcb = ycb = 5	xcb <= 6, ycb <= 6
Code-block style	SPcod, SPcoc = 0000 0000	SPcod, SPcoc = 0000 0000	SPcod, SPcoc = 00sp vtra where r = v = 0, and a, t, p, s = 0 or 1 NOTE — a = 1 for selective arithmetic coding bypass t = 1 for termination on each coding pass, p = 1 for predictive termination s = 1 for segmentation symbols
Transformation	9-7 irreversible filter	9-7 irreversible filter	No restriction
Precinct size	PPx = PPy = 7 for N_LLL band, else 8	PPx = PPy = 7 for N_LLL band, else 8	PPx >= xcb, PPy >= ycb
Tile-parts	Each compressed image shall have exactly 6 tile parts. Each of the first 3 tile parts shall contain all data necessary to decompress one 2K color component compatible to 2k digital cinema profile. Each of	Each compressed image shall have exactly 12 tile parts. Each of the first 3 tile parts shall contain all data necessary to decompress one 2K color component compatible to 2k digital cinema profile. Each of	Each compressed image shall have exactly Csiz tile parts. Each tile part shall contain all data from one component

	the next tile parts shall contain all additional data necessary to decompress the rest color component. The resulting codestream structure is diagramed in Figures A-29	the next 3 tile parts shall contain all additional data necessary to decompress one 4K color component. Each of the next 3 tile parts shall contain all additional data necessary for the rest of one 2k color component. Each of the next tile parts shall contain all additional data necessary to decompress one the rest of the color component. The resulting codestream structure is diagramed in Figures A-25, A-27 and A-28.	
Other markers			
Packed headers (PPM, PPT)	Disallowed	Disallowed	Disallowed
Tile-part lengths (TLM)	TLM marker segments are required in each image	TLM marker segments are required in each image	TLM marker segments are required in each image
Packet length, tile-part header (PLT)	For each tile-part a complete list of packet lengths shall be provided	For each tile-part a complete list of packet lengths shall be provided	For each tile-part a complete list of packet lengths shall be provided
QCD, QCC	Main header only	Main header only	Main header only
SOP, EPH	Disallowed	Disallowed	Each packet in any given tile-part shall be prepended with a SOP marker segment and each packet header in any given tile-part shall be postpended with an EPH marker segment
POC marker	There shall be exactly one POC marker segment in the main header. Other POC marker segments are disallowed. The POC marker segment shall specify exactly two progressions having the following parameters: First progression: RSpoc = 0, CSpoc = 0, LYEpoc = 1, REpoc = N_L+1, CEpoc = 3, Ppoc = 4 Second progression: RSpoc = 0, CSpoc = 0, LYEpoc = 2, REpoc = N_L+1, CEpoc = 3, Ppoc = 4	There shall be exactly one POC marker segment in the main header. Other POC marker segments are disallowed. The POC marker segment shall specify exactly four progressions having the following parameters: First progression: RSpoc = 0, CSpoc = 0, LYEpoc = 1, REpoc = N_L, CEpoc = 3, Ppoc = 4 Second progression: RSpoc = N_L, CSpoc = 0, LYEpoc = 1, REpoc = N_L+1, CEpoc = 3, Ppoc = 4 Third Progression: RSpoc = 0, CSpoc = 0, LYEpoc = 2, REpoc = N_L, CEpoc = 3, Ppoc = 4 Fourth Progression: RSpoc = N_L, CSpoc = 0, LYEpoc = 2, REpoc = N_L+1, CEpoc = 3, Ppoc = 4	Disallowed
Application specific restrictions			
Error protection	Disallowed	Disallowed	The use of marker segments defined in ITU-T Rec. T.810 | ISO/IEC 15444-11 for the detection, correction and protection against errors that may result from aging media is not mandatory but optional and strongly recommended.
Max compressed bytes for any image frame (aggregate of all 3 color components)	1302083 bytes	2604166 bytes	No restrictions
Max compressed bytes for any single color component of an image frame	1041666 bytes	2083332 bytes	No restrictions

Max compressed bytes for quality layer 0 of any image frame (aggregate of all 3 color components)	1302083 bytes for 24 fps 651041 bytes for 48 fps	1302083 bytes for 24 fps	No restrictions
Max compressed bytes for layer 0 of any single color component of an image frame	1041666 bytes for 24 fps 520833 bytes for 48 fps	1041666 bytes for 24 fps for 2K portion of each component.	No restrictions

Main header	Tile-part header	c0p*r*l1	Tile-part header	c1p*r*l1	Tile-part header	c2p*r*l1	Tile-part header	c0p*r*l2	Tile-part header	c0p*r*l2	Tile-part header	c0p*r*l2

Proposed codestream structure. Assuming N_L wavelet transform levels (N_L+1 resolutions), the rectangle labelled c*i*p*r*l1 (i = 0, 1, 2) contains all packets for color component i, all precincts, resolutions 0 through N_L and layer 1. The rectangle labelled c*i*p*r*l2 (i = 0, 1, 2) contains all packets for color component i, resolutions 0 through N_L and layer 2.

Digital Cinema and Object Oriented Sound Reproduction

Judith Liebetrau
Fraunhofer IDMT-Institut für Digitale Medien Technologie

Gabriel Gatzsche
Fraunhofer IDMT-Institut für Digitale Medien Technologie

***Abstract* - Within the EU-funded project EDCine strong effort is devoted to the improvement of the Digital Cinema movie experience. Due to the need to improve the sound and movie experience new approaches to generate and reproduce 3D sound scenes have been developed. In this context a new object-oriented audio representation is introduced and it is shown how such an innovative sound reproduction approach can be integrated into the Digital Cinema.**

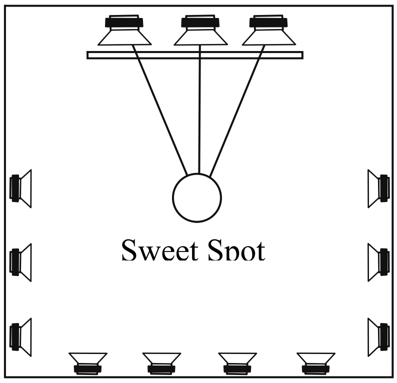
I. INTRODUCTION

The way of Digital Cinema sound reproduction did not change much in the recent years. Simply the number of loudspeaker channels and the audio quality of the signals feed to the speakers have been increased. While the new Digital Cinema specification allows up to 16 channels of uncompressed sound data, 5.1 Surround sound will continue to be the standard for the next years. An integration of additional loudspeakers is difficult and newer and more advanced reproduction approaches like Wave Field Synthesis (WFS) can't be pushed to the limits with the current state of the art. Therefore the sweet spot problem still exists: Only in the centre of the listening area an optimal sound can be provided (Figure 1). Furthermore in today's Digital Cinema sound systems a perceivable mismatch between the position of a visual object on the screen and the position of an audio object reproduced by the sound system can occur. Another issue of today's cinemas is the assumption of a well defined and correctly installed loudspeaker setup. If the loudspeakers are located at a position different from the position originally assumed, then the perceived audio quality is impaired.

II. STATE OF THE ART

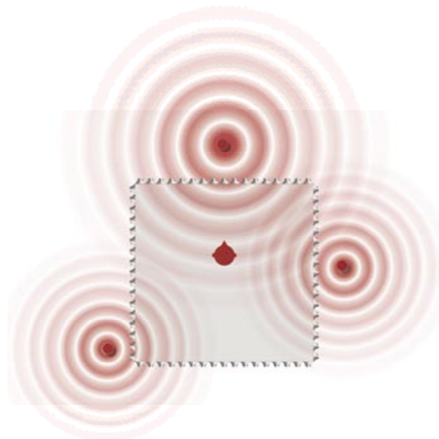
In cinema mono and 2-channel stereo formats for audio signals have been used since many years. In the 40's [8] first multichannel sound systems have been introduced to enhance the sound and movie experience. During the last years, further developments on multi-channel audio formats were done. Today the most common formats use five full range channels together with an optional enhancement channel for low frequency content (LFE). Also formats with four full range channels, six full range channels and seven full range channels are used. All these formats are based on the channel oriented paradigm: the layout of loudspeakers is taken into account in production.

FIGURE 1
COMMON LOUDSPEAKER LAYOUT



A sound reproduction technique which overcomes the restrictions of channel based formats is called Wave Field Synthesis (WFS). It was invented in the late 80^{th} by the Technical University of Delft (NL) [9]. WFS allows a very good localisation of the reproduced sound events for a large audience area.

FIGURE 2
WORKING PRINCIPLE OF WFS



WFS recreates the sound field (Figure 2, Red waves within the loudspeaker ring) of original audio sources (Figure 2, red points). This results in a very well distributed sound field and very stable sound source localization. In WFS production and storage of audio data usually follows an object oriented paradigm: an audio object consists of the audio content and its position in a scene. Due to technical and financial restrictions all current installations of WFS systems are restricted to perfect spatial reproduction in the horizontal plane.

III. Extensions to 3D

Within the EDCine Project the extension of cinema sound reproduction by a real third dimension was studied, this means the extension of the standard multi channel setup by additional loudspeakers at the ceiling and the floor was investigated [6].

Additionally, the usage of less dense loudspeaker setups i.e. the usage of fewer loudspeakers at positions above and beyond the audience's listening plane seems to be a first step towards 3D reproduction. A new 3D loudspeaker layout was developed as well as a driving algorithm. This algorithm works similarly to WFS in an object oriented way. Parametric descriptions of sound source properties like the position, source type, etc and additional control data are processed to derive a set of driving coefficients for the reproduction system's signal processing stages.

Furthermore during the EDCine project the possibilities of reproduction system independent spatial sound design tool were investigated. The aim was to provide sound design tools which abstract form the reproduction system. The results of such process can be adapted to multi-channel systems as well as to WFS reproduction [7].

IV. Object Oriented Sound Reproduction

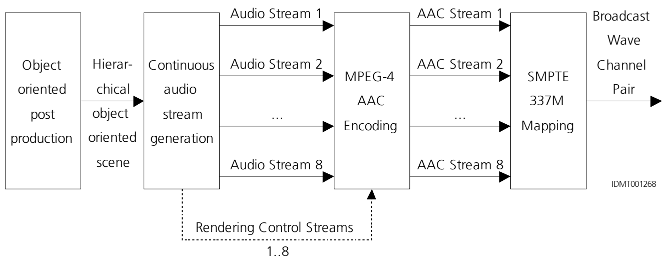
The channel oriented sound reproduction approach requires a well defined loudspeaker setup, i.e. the number and positions of the loudspeakers are predefined. The mastering process knows the target setup and prepares the loudspeaker signals in a way that they perfectly fit the assumed setup. This implies that it is difficult to feed the generated signals into another sound system. This problem is solved by the object oriented sound reproduction approach: Here a sound scene is provided in a conceptual way. The scene is composed of a number of sound sources that can be moved through the space. Sound sources can have directivity, spatial distribution and can interact with the virtual acoustic environment. Besides positioning of direct sound, a position dependent calculation of early reflections and diffuse reverberations is possible which makes it possible to generated realistic but also artificial spatial environments. Through the availability of the direct sound of each source and a parametric description of the properties the room the reproduction can optimal be adapted to the given spatial environment. This can be a Wave Field Synthesis setup but also an arbitrary loudspeaker configuration. Increasing the number of loudspeakers is increasing the size of the sweet spot and is making sound sources more stable. A movie theatre has much more freedom to decide which loudspeaker setup to be installed, because the actual loudspeaker signals are calculated at the reproduction side through a process called rendering. This approach allows also the rendering of additional scenes, for example for binaural signals for the VIP lounge. Also the emphasis of particular sound sources, e.g. the dialogs for hearing impaired is possible or the visualization of sound sources for deaf people can be done in a much better way. More details of the object oriented approach in Digital Cinema can be found in [1].

V. Integration of Object Oriented Sound into the Digital Cinema

A key question is how object oriented sound can be brought to the Digital Cinema. The state of the art, as for example defined by the Digital Cinema Initiatives (DCI), only allows the transmission of a small number of loudspeaker signals that are more or less directly fed to the installed loudspeakers. It is necessary to think about ways how to transmit an object oriented scene containing hundreds of sound sources using the existing Digital Cinema infrastructure. In [1] we proposed to limit the number of sound sources that are simultaneously active. This limit could be for example 64 simultaneous sound sources. With that step we guarantee that the data rate will not exceed a certain amount. The next step is to assign all sound sources to one of e.g. 64 audio tracks. The resulting streams are MPEG-4 AAC [3][4] encoded and mapped to a Broadcast Wave Channel pair using the SMPTE 337M Mapping [2] standard. The rendering control streams – that are the sound source properties like positions, directivity – are embedded into the AAC audio streams. This solution brings the synchronization of audio data and rendering control information. The resulting Broadcast Wave Channel pair can be now integrated into the standard Digital Cinema Content Package [5]. The proposed way has the following advantages [1]:

- Through the Broadcast Wave like data structure the OO-DCDM can be encrypted and packaged using the existing encryption and packaging tools.
- The MPEG-4 AAC data reduction makes it possible to convey 64 sound or more sound sources using the 16 standard Broadcast Wave audio channels.
- The mapping of the sound sources to 16 Broadcast Wave tracks additionally allows the usage of the existing Audio Media Block. The object oriented audio renderer can be directly connected via AES3 to the Audio Media Block.
- The embedding of the rendering control information directly into the appropriate AAC stream makes the implementation of additional synchronization techniques unnecessary.

FIGURE 3
THE EMBEDMENT OF 8 OR MORE SOUND SOURCES INTO ONE BROADCAST WAVE CHANNEL PAIR [1]



VI. TOWARDS OPTIMAL 3D SOUND: STOP THE BABYLONIAN CONFUSION

Object oriented sound will boost innovative reproduction approaches like WFS to its limits. Despite of this fact many different channel oriented sound reproduction do exist and will be available in the next years. A big problem is that all of these multi channel standards are more or less incompatible. The most common formats today use 5 full range channels together with an optional enhancement channel for low frequency content (LFE), but also formats with four, six and seven full range channels are used. A general problem for such systems is that even if the number of channels is the same different recordings/mixes might be based on a different layout of the loudspeakers. An example for six channels is 2+2+2 (two channels in front, two channels on side back, two channels elevated in front) which can easily be misinterpreted as 5.1 (three channels in front, two surround, one LFE). The way this problem is treated today is the definition of so called channel labels which assign the different audio signals of a multichannel stream to the appropriate loudspeakers. The difficulties are the following:

- The invention of a new loudspeaker arrangement requires the definition of additional channel labels. The standardization and management of new labels requires much effort. Furthermore the existence of many different labels is quite confusing.
- The channel labels do not imply the loudspeaker positions. If there are two multi channel audio tracks that have been produced for different loudspeaker arrangements it is not possible to adapt the signals dynamically.
- Realtime audio interfaces like MADI or ADAT do not convey the intended audio signal to speaker assignment. Therefore an additional data connection is required which conveys the intended assignment information.

To solve those difficulties, a data format is required which does not convey the channel labelling but the spatial location of the loudspeaker where the signal belongs to.

With this information a multi channel loudspeaker reproduction system make an optimal decision to which of the real existing loudspeakers the audio signal is fed. If there is not an optimal loudspeaker, a best possible mapping strategy can be found to map the audio signal to the actual speaker setup.

Such a data format would also solve the problem of the channel labelling and would open the usage of more flexible speaker setups.

VII. SUMMARY

The Digital Cinema specification of the Digital Cinema Initiatives (DCI) was the starting point of the EU project EDCine. Among others the following tasks have been subject of work:

- The extension of the existing multi channel loudspeaker setup by additional loudspeakers at the ceiling and the floor was investigated.
- The minimum number of loudspeakers that are required to reach a stable perception of sound source positions around the audience was investigated [6].
- A concept has been developed how to adapt the Digital Cinema production, transmission and reproduction chain such that object oriented sound scenes can be reproduced using as much as possible of the Digital Cinema infrastructure [1].
- A data format as well an encoder, decoder and a player have been developed, that allow it to package object oriented sound scenes into a Digital Cinema Content Package. The playout with an XDC Cinema Server and the reproduction with an IOSONO WFS system have been tested [1].
- The possibilities of reproduction system independent spatial sound design tool were investigated. The aim was to provide sound design tools which abstract form the reproduction system [7].

REFERENCES

[1] Gabriel Gatzsche, Benoit Michel, Jérôme Delvaux, Lutz Altmann, „Beyond DCI: The integration of object oriented 3D sound into the Digital Cinema“, NEM Summit 2008, Palais du Grand Large – Saint-Malo, France

[2] SMPTE 337M - Format for Non-PCM Audio and Data in an AES3 Serial Digital Audio Interface, SMPTE - The Society Of Motion Picture And Television Engineers. 2006. – Standard

[3] ISO/IEC 14496-3, MPEG-4 Audio, Information technology -- Coding of audio-visual objects -- Part 3: Audio

[4] ISO/IEC 13818-7 - Generic coding of moving pictures and associated audio information - Part 7: Advanced Audio Coding (AAC) / ISO/IEC – International, Organization for Standardization. 2004. – Standard

[5] Digital Cinema Initiatives, LLC, Digital Cinema System Specification - Version 1.2, March 07, 2008

[6] Judith Liebetrau, Thomas Sporer, Thomas Korn, Kristina Kunze, Christoph Man, Daniel Marquard, Timo Matheja, Stephan Mauer, Thomas Mayenfels, Robert Möller, Michael-Andreas Schnabel, Benjamin Slobbe, Andreas Überschär: “Localisation in Spatial Audio - From Wave Field Synthesis to 22.2”, 128th AES Convention, Amsterdam, 2008

[7] Frank Melchior and Diemer de Vries: Plane wave decomposition as a tool for interactive modification of measured sound field and their adaptation to different reproduction system, 19th International Congress on Acoustics, 2007
[8] WM. E. GARITY AND J. N. A. HAWKINS:"FANTASOUND", Journal of the Society of Motion Picture Engineers, 1941
[9] Berkhout, A.J.:"A holographic approach to acoustic control", Journal of the Audio Engineering Society, Vol.36, No.12, December 1988, pp.977-995.

Panel on New Paradigms for Creative Research

Organized by

Adaptive Architectures Group

Hogeschool voor de Kunsten Utrecht

Faculteit Kunst, Media & Technologie

Coordinated by

David Crombie, Utrecht School of the Arts (The Netherlands)

AXMEDIS 2008: proceedings of the 4th International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution. Workshops, Panels and Industrial Applications. Florence, Italy 17-19 November 2008
edited by Paolo Nesi, Kia Ng, Jaime Delgado
ISBN: 978-88-8453-811-6 (online). ISBN 978-88-8453-810-9 (print).

Outcomes from the EUAIN Network: improving access to books

Anne Bergman-Tahon
Federation of European Publishers
Brussels, Belgium
abergman@fep-fee.eu

David Crombie
EUAIN Network
Amsterdam, Netherlands
{dcrombie} euain.org

Abstract

The European Accessible Information Network (EUAIN) was established in 2004 when a core group of organisations involved in accessible content production came together on a European level to seek greater clarity and systematisation for this field. This was made possible through European Commission support under the 6th Framework Programme. This paper provides an overview of the outcomes from the EUAIN network including the recommendations on Article 6.4.1 of the EC Copyright Directive. These recommendations include a number of pointers for how better to make the majority of books published accessible/adaptable from the outset so that print impaired people have access to virtually all books when they are published.

1. Introduction

During the last 4 years the EUAIN Network [1] has brought together the different stakeholders in accessible content processing and sought to find new ways to mainstream the provision of accessible content. This has involved looking closely at production processes, supporting technologies, distribution and value-chain issues and new ways of meeting the core needs of print impaired people. It is becoming clear that a far deeper examination of fundamental accessibility is required if we are to mirror mainstream content provision. The EUAIN Network will act as a focal point to pursue these activities and to ensure that all European countries have access to appropriate training and expertise.

2. EUAIN Outcomes

An implicit result of EUAIN is the transparent integration of consumer and producer models for digital content. This integrated model will enable the inclusion of accessibility from the ground up, a key aspect of e-Accessibility and e-Inclusion. An additional key feature is the conformance of the standardisation process to well-known standardisation processes.

Through the provision of a competencies representation framework in the form of a network, designers, producers and consumers can interact. The framework will function as a communication vehicle between different actors and will provide a basis for future work. Moreover, it will enable the involvement of end-users into the design trajectory. Through this, the end-users are likely to develop a sense of commitment to the product development and the partners involved in the product development. This will help to raise the public acceptance of the e-Accessibility venture and in the end raise the demand for accessible content and associated products. The e-Accessibility representation network and its knowledge base with its knowledge distribution infrastructure will ensure availability of the knowledge and technology harvested through this collaboration.

3. CEN Workshop on Document Processing for Accessibility (WS/DPA)

Given the widespread adoption of ICT within the publishing industries, there is a general interest in the creation and provision of well-formatted digital documents. For those people who are dependent on accessible information, this interest is of central importance, and it is this convergence of interests that stimulated the creation of this Workshop. The CEN WS/DPA [2] has examined some of the ways in which this convergence is helping to build consensus and create new strategies and technologies for the provision of information in formats that are more accessible for everyone.

In the real world, publishers rely on accessibility experts and consider accessible information only at the end of the content production chain. This requires considerable amount of efforts to make information accessible for everyone and it is a very hard problem to tackle. This workshop introduces accessibility as a design element in content production and provides guidelines and best practice on how more accessible documents can be produced. Another important issue is that the user requirements for accessible information are not well defined. In this work, we therefore based the elaboration on publishing use cases and scenarios that have been derived together with publishers in order to analyse at least partly the user requirements. Additionally those scenarios provide specific examples of accessible information provision as an entry point to publishing stakeholders.

The WS/DPA Workshop brought together some of the key players working in the fields of publishing and accessibility. The topics addressed ranged from generic document and knowledge structures, through all aspects of accessible document processing, to Digital Rights Management and copyright issues. Perhaps the most striking aspect was the level of convergence between the needs of accessibility communities and those of content creators and providers. Indeed, with the introduction of accessibility from scratch, the information needs of all consumers are better served, particularly as content providers seek innovative solutions for re-aggregating their content for new marketplaces.

The CEN DPA Workshop as detailed in its business plan had the following objectives:

• To bring together all the players in the information provision and e-publishing chain in order to achieve the critical mass significantly to enhance the provision of accessible information at a European level.

• To provide guidelines needed on integrating accessibility approaches and workflows within the document management and publishing process rather than as just a specialised, additional service.

• To raise awareness and stimulate the adoption at local, regional, national and European levels of the emerging formats and standards for the provision of accessible information and to find ways of ensuring that technological protection measures do not inadvertently impede legitimate access to information by people with print impairments.

Based on those objectives the workshop document:

• describes the outcomes from the DPA Workshop activities

• provides an elaboration of relevant standards and their possible use in the publishing sector

• examines the different formats required for accessible information provision

• provides a systematic overview of relevant conversion processes and related structured information activities

• examines possible scenarios of use within the publishing sector

• provides real-life case studies and instances of best practice

• identifies areas for further research and systematisation

The Workshop was initiated and supported by the EUAIN Network and is available from the CEN website [2].

This CWA addresses accessibility in the publishing value chain and examines ways to introduce and enhance accessibility of publishing content inside publishing workflows. The intended audience includes actors and stakeholders within the publishing value chain (publishers, authors, content producers and distributors, publishing system developers and vendors) and the content accessibility area (specialised libraries, accessibility consultants, and accessible system developers and vendors).

The CWA aims to provide a first elaboration on how the accessibility of publishing content can be enhanced by altering existing publishing workflows and introducing accessibility considerations where appropriate. For reaching this goal, in each step where accessibility is introduced, relevant formats and conversions are detailed out, as well new workflow items described.

Accessible information is not a special type of information aimed at a specific group of a certain population. *Accessible information* is information that can be accessed by anyone, with or without a disability, aimed at a general market where anyone interested is a possible customer. Structured information is the first step in the accessible information process. A document whose internal structure can be defined and its elements isolated and classified, without losing sight of the overall structure of the information, is a document that can be navigated.

Most adaptive technology allows the user to access a document, and to read it following the "outer" structure of the original. But if the same information has also an "inner" structure that allows the adaptive device to distinguish between a phrase and a measure, between a paragraph and a sentence, highlighting particular annotations, then the level of accessibility (and therefore usability) of the whole document will be greatly enhanced, allowing the user to move through it in the same way as those without impairments do when looking at a printed document, and following the same integral *logic*.

In an ideal world, all documents made available in electronic formats should contain that internal structure that benefits everyone. Highly-structured documents are becoming more and more popular due to reasons that very seldom pertain to making it accessible to persons with disabilities. The move to XML related formats and associated standards for metadata have provided an impetus for far greater document structuring than before. Whatever the reasons behind those decisions are, the use of highly-structured information is of great benefit to anybody accessing them for any purpose.

In recent years, the market for accessibility and assistive technologies has started to gain recognition. It is clear that the integration of accessibility notions into mainstream technologies would provide previously unavailable opportunities in the provision of accessible multimedia information systems. It would open up modern information services and provide them to all types and levels of users, in both the software and the hardware domain. Additionally, new consumption and production devices and environments can be addressed from such platforms and this would provide very useful information provision opportunities indeed, such as information on mobile devices with additional speech assistance. It is equally clear that we remain at the very beginning of the move to incorporate accessibility within mainstream content processing environments.

4. Recommendations to the EC on article 6.4.1 of the Copyright Directive

A key task of the EUAIN Network was to examine the extent to which the provisions of Article 6.4.1 of Directive 29/2001/EC (the European Copyright Directive) [3] have been effective. Article 6.4.1 reads:

“Notwithstanding the legal protection provided for in paragraph 1, in the absence of voluntary measures taken by rightholders, including agreements between rightholders and other parties concerned, Member States shall take appropriate measures to ensure that rightholders make available to the beneficiary of an exception or limitation provided for in national law in accordance with (...) (3)(b) (...) the means of benefiting from that exception or limitation, to the extent necessary to benefit from that exception or limitation and where that beneficiary has legal access to the protected work or subject-matter concerned.”

The aim of that Article is to ensure that the rights granted by copyright exceptions to (inter alia) people with reading related disabilities are not negated by technological protection measures (TPM).

To this end, the intended goal of the Directive is best described in recital (43): *“It is in any case important for the Member States to adopt all necessary measures to facilitate access to works by persons suffering from a disability which constitutes an obstacle to the use of the works themselves, and to pay particular attention to accessible formats”*. This is to be completed by the limitation itself (Article 5.2.(b)): *“...uses, for the benefit of people with a disability, which are directly related to the disability and of a non-commercial nature, to the extent required by the specific disability”.*

The report concludes that access problems undoubtedly exist, but that it is too early to draw firm conclusions on the effectiveness of Article 6.4.1. No relevant case law has been identified, and indeed the provisions made by most Member States in transposing of the Article are not well known even to those who might benefit from them. Furthermore, the effect of the Article 6.4.1 is likewise untested.

Based on the final outcome of the deliverable, the following recommendations are made to the European Commission and to other stakeholders.

The issues that have been identified can be divided into two main categories:

1. Access to printed works
2. Access to electronic works

4.1 Access to printed works

For granting permission to scan the actual book to produce an accessible format there are two solutions. The first is to develop voluntary agreements between rights holders (publishers) and institutions representing/serving reading disabled persons. The second is to designate one trusted repository where publishers can deposit their books and which will serve as a hub for institutions representing/serving reading disabled persons.

For providing the electronic file that was used for printing the book there are two solutions. The first is to develop voluntary agreements between rights holders (publishers) and institutions representing/serving reading disabled persons. The second is to designate one trusted repository where publishers can deposit their books and which will serve as a hub for institutions representing/serving reading disabled persons.

4.2 Access to electronic works

To provide access to non-accessible/adaptable electronic books protected by Technical Protection Measures (TPM) there are two possible solutions. The first is to develop voluntary agreements between rights holders (publishers) and institutions representing/serving reading disabled persons. The second is to designate one trusted repository where publishers can deposit their books and which will serve as a hub for institutions representing/serving reading disabled persons.

To provide access to accessible/adaptable electronic books protected by TPMs there are three solutions. This is for accessible/adaptable electronic books protected by TPMs which prevent reading disabled persons to 'read' the book. The first solution is to encourage publishers to label these books. The second is to work with publishers to see how TPMs impede the access for reading disabled persons and to change these features of the TPMs. The third is to designate one trusted repository where publishers can deposit their books and which will serve as a hub for institutions representing/serving reading disabled persons.

To provide accessible/adaptable electronic books protected by TPMs which do not prevent reading disabled persons to 'read' the book there is one solution, namely to properly label the work and put the necessary information on the web site so all potential users of the book know what they are or are not permitted to do.

As the copyright exceptions (5.3.b.) have been implemented in all EU countries and though while acting as a strong incentive to find consensual solutions, they are not seen as the ultimate solution, *it is proposed that the Commission develops guidelines* to encourage rights holders and institutions representing reading disabled persons to find the best ways to make most of the information accessible to all.

4.3 Practical Suggestions

These guidelines should be based on the following principles:

- To achieve what should be the common goal of both publishers and institutions representing disabled persons, supported by the EU institutions and their respective Member States, the collaboration between all stakeholders is an absolute key issue and can be enhanced.
- In the absence of accessible/adaptable books, publishers should be encouraged to make their content accessible through trusted third parties (ideally one per country, maybe more when languages are not homogenous within a country). To this end, they should either permit the trusted third party to digitise the book and make it available, against remuneration if jointly agreed, to reading disabled persons within extranets, or they should be providing the electronic file which has been used by the printer, to facilitate access to reading disabled persons within extranets.
- The same applies to electronic books which would be published in a non-accessible/adaptable version; the publishers should be encouraged to provide the electronic file to a trusted third party, which in turn will provide access to reading disabled persons within extranets.

But the real goal, likely to be achievable in the digital world, is to make the majority of books published accessible/adaptable from the outset so that reading disabled persons have access to virtually all books when they are published. They will no longer need to access them through institutions serving reading disabled persons but will directly acquire them through online retailers or high street booksellers. It is foreseen that libraries will retain an important role in the digital world and that as recommended in the 2001 Copyright in the information society Directive in recital 40: *"Therefore, specific contracts or licences should be promoted which, without creating imbalances, favour such establishments and the disseminative purposes they serve".*

How can the EU practically support such a goal? We offer the following suggestions:

1. Encourage publishers and expert bodies to pursue their fruitful dialogue.

2. Support the work of the EUAIN Network to be established as an autonomous, not-for-profit foundation that can build upon this solid dialogue.
3. Involve the software developers in the dialogue encouraging to propose the development of publishing software better adapted to the needs of reading disabled persons and delivering high performance workflows for the publishing industry.
4. Work on common European standards for conversion software that could then be used by publishers, whether or not in connection with TPMs or rather Digital Rights Management (which will allow an indefinite number of business models inclusive of all users).
5. Encourage the publishing industry to work closely with expert bodies to ensure that all accessibility guidelines in the design of digital material are followed as a matter of course. This includes for example encouraging publishers to properly tag their books so that they can be accessed by all without third party intervention (it is understood that this will not apply to some books which are too complex for publishers and will always require the intervention of a specialist agency).
6. Encourage the Member States to designate one trusted third party, when not yet in existence, to which the publishers could provide their books or, even better, electronic files upon request, to be adapted/made accessible for reading disabled persons. In agreements between parties, we recommend that priority be given to developing technical solutions such as; provision of an encryption key to the trusted third party; developing watermarking and fingerprinting techniques; creating extranets such as web sites accessible only to authorised people, where access could be tailored to individual users' needs.
7. In addition, these trusted third parties could serve as partners in drawing comprehensive and straightforward voluntary agreements at national level (eventually with some input of the EU) to facilitate the prompt resolution of any TPM-related access difficulties that may from time to time arise.
8. The Commission could set the example in preparing guidelines to help all parties (publishers, consumers, legislators and the judiciary in each member state) to determine the best way of resolving conflicts that may arise. Such guidelines should seek to reduce national differences.
9. Develop services such as Publisher LookUp, developed by the Association of American Publishers. This facility designates an individual in each publishing company who can deal with requests for access from people unable to access the standard version of a work. This must be done bearing in mind the size of publishing companies in Europe.
10. In the absence of accessible/adaptable version and in the case of TPMs preventing use of conversion software, labelling schemes for products endowed with DRM should be developed. Any labelling scheme should be used to indicate clearly how the bona fide beneficiary of an exception can gain ready access to the material in question, whether that is through the publisher or through technological means.
11. Keep monitoring the issue to eventually review the situation also in view of the development of the digital market place. This will need to be backed up by further research, including not only surveys but also activities such as workshops with all stakeholders. The Commission should consider series of workshops around Europe to increase awareness and understanding and promote best practice.
12. Develop legislation on taxation issues to provide an incentive to publish an increasing number of books accessible to reading disabled persons (such as the possible technical adaptation of Annex 3 of the 6th VAT Directive.

5. Collaborative Working

By raising awareness of the issues discussed above and by bringing together the key actors at a European level, EUAIN provides a knowledge base that can be accessed by all those involved in content creation and consumption, which is to say both producers and consumers. This is an ambitious goal, but the convergence described above makes this both worthwhile and achievable.

One of the main objectives of EUAIN project was the creation of a collaborative working knowledge and expertise environment in both a traditional and new economy sense. The newly established EUAIN Network seeks to marry technological knowledge in the domain of accessibility in the broadest sense with fundamental scientific knowledge in the domain of accessibility. An additional objective is the distillation of an educational framework that will ‘engrain’ notions about accessibility (captured in the terms e-Inclusion and e-Accessibility) into modern society. The EUAIN Network aims to explicitly address this task at a fundamental level by addressing the provision of accessible information for all EU citizens in member states. It is our belief that this activity can only be successful when addressed at a European level and when fully integrated with existing initiatives in the accessibility and learning standardisation processes.

Notions of "accessibility" are normally equated with the adaptation and conversion of digital content, where this content can be made available. On a European level, and indeed often on a national level, much of the existing expertise on creating accessible adaptations of digital content is of a highly distributed nature. Within specialist organisations supporting print impaired people; or within university research laboratories; or indeed within publishing houses, many automated tools have been designed and implemented at least partially to execute the necessary adaptation procedures. However, each automated tool has its own, highly specific, field of application. Furthermore, the knowledge required to build these very specific tools is equally distributed, so that there is currently very little re-use of either tools or knowledge. Indeed, the approach taken by the EUAIN Network was acknowledged in the recent World Intellectual Property Organisation (WIPO) report[4]:

"Built in access for visually impaired people right from the start does, nevertheless, seem to be a highly desirable way forward, but stakeholders need to be aware of the problems due to lack of standards, ever-changing technology, use of DRMs and so on, as well as possible solutions, in order to ensure built in accessibility is not just a theoretical solution. In this respect, the work of EUAIN which, as already mentioned, is described in case study 13 of Chapter 5, brings together a range of stakeholders to explore issues such as these. This is perhaps an example of a way forward more generally and work of this nature should perhaps be promoted more widely by governments and international agencies. It seems to be in everyone's interests that a desire to build in access from the start is both encouraged and facilitated by ensuring that what this requires in practice is widely understood and adopted".

6. Future Research Areas

It is becoming clear that a far deeper examination of fundamental accessibility is required if we are to mirror mainstream content provision. Given the general move towards distributed media, there is a need to develop open source frameworks to bridge the gap between original content design heuristics and intuitive multimodal interfaces required for content and communication systems. Such frameworks would provide built-in, profile-based access to information, content and services, which not only combine and extend state-of-the-art technologies for information access, but also conform to standards and guidelines available for accessibility, usability, scalability and adaptability.

Individual actors in the information provision chain cannot tackle the coherent and sustainable provision of accessible information in isolation. While examples of good practice are emerging in the production sphere and in new collaborative distribution models, a European wide approach offers far greater potential. By examining the key areas at a European level, the emerging knowledge in this area will be more widely applicable and far more likely to stimulate new initiatives and research for producing and distributing good practice in accessible information.

As noted above, previous and ongoing work in this area has been sporadic, and there has been little attention paid to the profound need to disseminate clear and practical advice to the different actors in the information provision chain. The EUAIN Network recognises that there is a clear need to fill this vacuum with unambiguous and clear guidelines, recommendations and standards from within the context of an integrated processing framework.

This will be done in such a way as to ensure widespread awareness raising and dissemination within the information provision industry and to the general public. Additionally, much of the knowledge gained by the work undertaken will be collated, examined and codified into new standards for accessible content processing. This process involves the creation of open standards, and the process will in itself help to raise awareness of key issues and stimulate discussion.

7. References

[1]**http://www.euain.org**
[2]**http://www.cen.eu/cenorm/businessdomains/businessdomains/isss/activity/ws-dpa.asp**
[3]**http://ec.europa.eu/internal_market/copyright/documents/documents_en.htm**
[4]**http://www.wipo.int/meetings/en/doc_details.jsp?doc_id=75696**

The ProAccess project: Improving Accessibility of Educational Content

Maria Loi & Claudio Li Calzi
Italian Publishers Association (AIE)
Milan, Italy
{maria.loi}{Claudio.licalzi}@aie.it

Ted van der Togt & David Crombie
DEDICON
Amsterdam, Netherlands
{tvdtogt}{dcrombie}@dedicon.nl

Abstract

Improving accessibility of educational material for visually impaired people is the main pillar of ProAccess [1] project. It aims at providing publishers and intermediaries in the elearning value chain (libraries, schools, charities and associations devoted to impaired people) with practical guidelines and instruments for the production and use of accessible content in a more effective way both from the productive process and copyright standpoint. The primary goal of the ProAccess project is to improve accessibility of educational content in the eLearning value chain for visually impaired people. The project disseminates the best practices and guidelines stemming from the results of two EU funded projects: ORMEE [2] and EUAIN [3].

1. Introduction

In recent years, specific sets of legislation, e.g. the UK Disability Discrimination Act [4] ("Reasonable adjustment and Special Education Needs Discrimination Act") or the so called Legge Stanca [5] in Italy are either stimulating or forcing content providers to accelerate the integration of disabled people, giving them the opportunity to access the educational materials in the different format needed to be used in efficient way. At European and national level it has been emerging major awareness and social responsibility amid educational content producers on the importance to provide accessible materials timely, allowing disabled people to get the same materials as their schoolmates and avoiding any social divide. As a matter of fact, impaired students still receive their course materials with a big delay compared with their schoolmates and this situation may cause them difficulties in regularly attending the school courses. Therefore, this is a topical issue which must be tackled as content producers encounter major difficulties specifically in managing the workflow involved in the preparation of the accessible versions of the schoolbook such as Braille, large fonts and digital version, being textbooks and other didactical materials (e.g. supplementary materials or digital content as Learning Objects) indeed one of the most complex kind of published material in terms of structure, graphic and layout.

2. ProAccess Approach

The main objectives of the project were achieved by following these steps:

* By evaluating the current situation in the involved countries, the needs of the disabled people and the problems arising from their requests to the publishing sector will be analysed, involving in the process key schoolbook publishers and representatives of disabled people

* Starting from the results of the EUAIN project, the production process required to produce accessible documents will be defined

* By analyzing the content value chain in the education sector, a set of shared rules in managing rights will be set out starting from the conclusion of the ORMEE project

* Promoting results in the publishing industry and within intermediaries and disabled organisations will foster adoption of publishing workflow that considers accessibility from scratch and a correct management of copyright.

The main value added of the project is to be found in its collaborative approach as the proposal arises from the awareness that proactive rights management and collaboration between different stakeholders may be effective tools to improve accessibility and broadly to increase access to digital content.

3. ProAccess Outcomes

The idea of grouping publishers and conversion companies in order to test the conversion processes required for different kinds of content stems from the achievements of the Pro Access workshop held in Venice on January 2008 [6] which gathered representatives of publishing industries, blind organisations and experts in conversion activities. As a result, representatives of three major Italian publishing houses participating at the workshop agreed to contribute at the testing phase with experts of conversion activities. Therefore, they provided a sample of their own production in order to shed light on the different issues in obtaining accessible versions for a school textbook, a novel or a dictionary – just to mention some kinds of content.

3.1 Overview of issues in converting publishers' output files into accessible versions for users with print impairments

Pro Access is a European Commission funded project aimed at enhancing accessibility of content by fostering linkages between the publishing industry and specialist organisations - be they universities, companies or associations - committed to providing content suitable for impaired users. During the course of the project, the consortium noticed the necessity of assessing the production constraints that currently prevent impaired users from having access to content in different versions other than paper. As a matter of fact, current processes are primarily designed to get the traditional paper, so that additional efforts are needed to attain accessible versions of the same output. This understanding has led to the creation of a working group coordinated by the Associazione Italiana Editori (AIE) [7] that brought together both representatives of the Italian publishing industry and organisations at EU level skilled in converting content into accessible versions.

The underlying rationale was to build a comprehensive group of actors which might illustrate the current restraints in creating accessible content both from the point of view of the publishers and specialist organisations, whose efforts are counted towards providing users with content in Braille, Daisy or large print version. It indeed emerged during the project meetings held with publishing companies and actors skilled in conversion activities that these industries are currently not working, communicating and exchanging knowledge, with the ensuing result that most of the actual conversion activities entail lots of manual processing and duplicate efforts which indeed make them costly and inefficient.

As the most difficult typology of content to be converted emerged to be that with particular elements such as tables, images, text with formulas and symbols, graphs – namely content designed for primary and secondary education pupils – the testing phase was oriented towards the most prominent Italian publishers whose core business was centred on the educational content market.

Because of the Italian Publishers Association's strong ties with the most distinguished publishing companies entrenched in the education, dictionary, fiction and non-fiction segment of the market, it has been agreed to address the top management of three major companies. These have been chosen on the basis of their huge production of educational books, thus encompassing all the different typology of content – ranging from math, history and geography textbooks to novels or reference books.

The conversion testing phase aimed at highlighting the steps, the choices to be made as well as constraints in converting books provided by publishers into accessible versions, underlining the specific workflow as well as all the different aspects and issues to be tackled in producing output material for impaired users.

As a matter of fact, assessing current publishing workflows and proposing innovative paths to make content accessible have been the primary focus of the current analysis, thus allowing to envision alternative production processes which mainstream accessibility features.

3.1.1 Objectives

The objectives of the conversion test were the following:

- Putting together both publishers and specialist organisations skilled in conversion activities in order to underline how more efficient solutions within the current publishing workflow might be attained to deliver content more easily convertible for users with certain reading impairments
- Enhancing communication and integration among the different actors involved in the conversion activities, namely publishers and specialist organisation – the latter being often little aware of the different steps required within the publishing workflow for getting the output

- Proposing feasible solutions to address efficiently the issues of converting different content's typologies and elements
- Raising publishers' awareness that producing accessible content might broaden market potential

3.1.2 Testing phase

The publishers committed to providing the material to be used for the conversion activities represent a sample of the most distinguished companies of the industry, covering a remarkable portion of the Italian publishing market other than a wide range of different content complexities.

The top management of the publishing houses have been asked to collaborate in the testing phase by means of providing samples of their publishing production intended to cover all the different typology of content and complexities such as text with tables, graphs, images, particular graphic elements and so on.

The underlying rationale was to get a comprehensive set of different documents that might suffice to show all the different and specific issues in converting content into accessible versions.

Having obtained the endorsement to undertake the test by the publishers' top management, AIE began to work in strict cooperation with representatives in charge of internal book production processes in order to agree on the suitable formats for the conversion-testing phase to be handed over to specialist organisations and accessibility experts. These were chosen on the basis of their recognized expertise in addressing different accessibility issues according to their specific working field.

AIE was in charge of coordinating the deliver of the output files to conversion specialists in the following formats:

- Print-ready Pdf – commonly used for delivering the book to printmakers
- Quark Xpress [8] or Adobe InDesign [9] if available – the most widely used page-layout and design software developed for the publishing industry.

The rationale for choosing the cited formats has been to highlight the extent to which the current publishing outputs actually smooth the way for subsequent work of conversion specialists or the other way round.

According to each specific working field and expertise, publishers' files have been submitted to conversion specialists through the intermediation of AIE both in Quark Xpress (or Adobe InDesign if available) and print-ready pdf format.

The content provided covered a wide range of different types of text, from the simple textbook as could be the case of a novel, to more complex ones such as a science textbook with tables, graphs, captions, compound layouts, math's text with symbols and formulas, reference books as dictionaries, school titles like history schoolbooks plenty of images, maps, tables etc.

Such a diversified list of content was lumped together to shed light on the different conversion activities as well as the constraints to be addressed on making the content accessible in each case, thus covering the majority of content production's typologies in the educational market.

Each expert involved in the testing phase was asked to compile an evaluation sheet after the conversion activity previously agreed with the project partners in order to standardize the assessment process. The evaluation sheet encompassed the types of accessible formats produced, tools used for converting content other than asking time needed on average to make the accessible version and the issues encountered, if any.

The experts concluded with drawing attention to the tools used for conversion activities, the time effort spent for making the accessible version, assessing also the accessibility of the content delivered by publishers and the steps undertaken to get the accessible output.

3.1.3 Main Findings

The following paragraphs highlight the degree of accessibility of the content provided by the publishers and the ensuing attempts to convert it into accessible version according to the type of content:

- Books with symbols (e.g. school textbook for ancient Greek teaching)
- Books with images, captions, tables, graphs (e.g. Science, History, Geography schoolbooks)
- Books with formulas and symbols (e.g. Maths textbooks)

It has been indeed detected that content typology deeply affects subsequent conversion workflows. As a matter of fact, different conversion issues have been found according to the type of book, so that it is crucial to focus on the different elements of content to be converted – be them plain text, tables, images or graphs - in order to understand how each specific issue can be tackled.

The effort needed to convert the source file proved to be highly different according to the complexity of the content meaning that the conversion processes involved entails several actions currently far from being time and cost efficient. If such effort in time were translated into actual outlays, the expenditure for making content accessible would be currently unbearable both for the specialist organisation sustaining the production costs and for the final users which would be charged higher prices for getting the accessible version – unless its production is subsidized by third parties, namely State-funded organisations, Government's Departments or international institutions such the European Commission.

Finally, the following findings for each book's typology have been structured in order to outline these points:

- Description of the book features in terms of specific elements such as tables, graphs, maps, compound images with text attached, formulas, music scores and so forth
- Conversion issues encountered according to each specific element above mentioned
- Suggestions and recommendations for publishing houses on producing accessible content
- Evaluation of the average conversion effort needed in order to reckon the efficiency of the conversion process

4. Recommendations

Current conversion processes used by accessibility organizations are based on lots of manual processing. It would be a major step forward if publishers could deliver content in formats that are easy to convert into accessible formats like DAISY [10] or Braille. Tests were performed to see whether it is possible to convert existing (unaltered) publisher documents (in Quark Xpress/Adobe InDesign) to a XML based transfer format, which could be better used as starting point, than the current formats (like PDF). If publishers could do such conversion themselves in the future, this would dramatically improve the efficiency of the overall process. Note that it would not take away the need for accessibility organizations, which still have to do description of pictures, narrations of audio etc. , but much of the monotonous retyping, restructuring, could be avoided.

There are two main criteria for producing actual accessible content: Structure and description of visual elements. For structure it would help if the textual content could be delivered in a proper XML-based file format, in which different headings would be tagged correctly and elements put in a proper reading order.

Other than the structural elements, there will be visual elements in a document that somehow have to be 'translated' into something that makes sense to a visually impaired user. For this 'translation' experts will be nevertheless needed, since describing (all elements of) a picture is one thing, but conveying the educational concept behind it is something else.

From the main findings of this project the following recommended actions could be formulated:

- Putting together both publishers and specialist organisations skilled in conversion activities. In order to underline how more efficient solutions within the current publishing workflow might be attained to deliver content more easily convertible for users with certain reading impairments.
- Enhancing communication and integration among the different actors involved in the conversion activities, namely publishers and specialist organisation – the latter being often little aware of the different steps required within the publishing workflow for getting the output.
- Proposing feasible solutions to address efficiently the issues of converting different content's typologies and elements.
- Raising publishers' awareness that producing accessible content might broaden market potential.

The following issues should also be taken into account:

- Types and steps of the conversion process undertaken.
- Types of accessible outputs provided.
- Time and costs involved according to the different accessible format produced currently not efficient.
- Evaluate how the conversion activity could deliver their outputs more efficiently through the provision of better-structured files from the publisher.

5. References

[1]http://proaccess.euain.org
[2]http://www.elearningeuropa.info/directory/index.php?page=doc&doc_id=4727&doclng=6
[3]http://www.euain.org
[4]http://www.statutelaw.gov.uk/content.aspx?activeTextDocId=3330327
[5]http://www.pubbliaccesso.gov.it/normative/legge_20040109_n4.htm
[6]Proceedings of the workshop can be downloaded from the website http://proaccess.euain.org/
[7]Italian Trade Association of Book Publishers – http://www.aie.it
[8]http://www.quark.com/
[9]http://www.adobe.com/products/indesign
[10]http://www.daisy.org

Gutenberg 2.0: Towards Adaptive Publishing Environments

Maria Loi
Italian Publishers Association (AIE)
Milan, Italy
maria.loi@aie.it

David Crombie, George Ioannidis & Reinhard Ruemer
EUAIN Network
Amsterdam, Netherlands
{dcrombie}{gioannidis}{rruemer}@euain.org

Abstract

European publishers are global leaders in their field and the book sector in Europe currently has a retail turnover of about €40 billion per year. Books are obviously the fundamental vehicle of European culture, knowledge and languages, which the European Union (EU) seeks to promote. However, unlike the US, at a European level there is an absence of a coherent research policy covering programmes adapted to this sector, whether for intra-community projects or for those concerning the dissemination of European books outside the EU, a particularly noticeable gap given that 2009 is European Year of Creativity & Innovation.

1. Introduction

The exists a need to provide a sound basis for research and training for this sector and to help create an integrated roadmap to elucidate adaptive architectures for the design, processing and consumption of the next generation of digital content. Whilst the publishing sector makes extensive use of modern ICT, the changes in the culture of information usage are driven by other players and that the publishing sector has been rather slow [1] in providing adaptive content for emerging new devices, styles, methods and business models. A research focus on (meta)adaptivity seeks to redress this imbalance. It is no longer only the user who must adapt to the media and modalities; the media and modalities must themselves adapt to the needs and wishes of users.

The information culture is changing and this requires corresponding adaptations within the publishing sector, especially in the training curricula of young researchers. There is a need to bring together key players and content providers in the multimedia publishing industry, the creative industries and those involved in information design and provide both theoretical and practical research opportunities. This would provide a focal point for discussing our emerging information design and consumption needs and provide cross-disciplinary roadmaps and curricula which can be used within both traditional and newly emerging industry value chains. There has been much discussion on the future of the book in recent years but *almost no attention to the implications of information convergence from the design perspective*. There is still a need to bring together these fragmented initiatives and provide a sustainable roadmap, network and PhD programme for future work in this area to support the training of young researchers and to assist the publishing sector to develop new knowledge and practice.

If we focus information representation theories, innovative computer science resources and contemporary insights onto the potential market of fundamentally adaptive information processing in shared frameworks. By structuring these key components on a meta-level that balances sufficient domain (inter)dependence and interoperability, solutions to existing and new problems and requirements may be achieved. In this way, existing approaches may be used to address contemporary adaptivity problems and provide a shorter time-to-market trajectory.

The motivation for establishing such a network to support these activities is borne of extensive experience at a European level on key requirements for industrial training and research for this field. The work is partially based on the existing EUAIN Network (European Accessible Information Network) and brings together key players from industry, administration and research. EUAIN has identified the key challenges of the publishing sector related to technical, organisational, socio-economical and political/legal as well as domains asking for in-depth research activities. The challenges of the existing state of the art are to be addressed in the following themes:

- Adaptive Architectures
- Adaptive Content Processing
- Adaptive Interface Design

In order to provide an industrial context for this work, and through close collaboration, exchange and training between the network participants and the associated partners, the practical socio-economic, legal and business aspects will then be considered within thematic adaptive environments. With the active core involvement of the Italian Publishers Association (AIE) as a network participant, and the Federation of European Publishers (FEP) as a key Associated Partner alongside a further seven associated industrial partners, we expect to achieve significant pan-European impact.

2. Objectives

The **GUTENBERG 2.0** training network has the following primary goals:

- to research and analyse recent developments from disciplines such as computer science, communication theory and interaction design to wider environments with a special focus on industrial and business applications
- to provide a solid framework for greater contact between academics and students with commercial research and development departments within a set of pre-defined parameters
- to provide coherent strategic output for the pan-European publishing industry upon which greater interaction between academia and industry can be fostered, with significant new career opportunities

In order to achieve these goals, the network will support ***both basic and applied research***. The **GUTENBERG 2.0** network will also provide outcomes upon which further work can be based. In particular, it will:

- allow interaction with the mappings between users and content from a higher perspective, thereby addressing a core industry concern
- annotate bottlenecks and optimal processes in the user to content communication path
- adapt content presentation according to the mapping descriptions between a user or group of users and the digital content
- enhance the detail of description of the user's experience dynamically, providing more parameters for optimisation of the communication process
- cross-link (basic and/or semantic) descriptions originating from users or user groups that reside in different processing stages. i.e. content creators, (re)producers, archivers, distributors, educators and consumers

Especially because **GUTENBERG 2.0** aims to design systems for all, it is essential to apply an open user centric and co-creative user involvement methodology in order to capture changing user requirements and deliver true innovative solutions that allow interfacing between the user models on one hand and the content and content semantics models on the other hand, that determine the level of flexibility and freedom these users have in choosing themselves how they want to explore the content. In this way a central concern can be addressed, namely the real-life mainstreaming of adaptive content processing within existing and emerging service provision and value chains.

3. Key areas

In order to ensure that universal access is a pre-requisite for future software, it is necessary to approach adaptive system design issues at a fundamental level. Unfortunately many solutions within mainstream environments exist as afterthoughts that are "*piggybacked*" onto the original design. It may seem obvious, but an adaptive framework which does not interact in any way with the core system architecture cannot use the design goals of the original system and therefore does not have the necessary integration in order to meet stringent sets of user requirements.

GUTENBERG 2.0 will investigate the phenomenon of adaptivity as well as conceive systems that allow the development of tools and interfaces that will aid designers *to focus on the adaptive behaviour of systems by enabling system modeling on a meta-adaptivity level.* The **GUTENBERG 2.0** network aims to investigate the preservation of resources and their inter-relations in a sustainable manner, by means of preserving the accessibility of the micro-, meso- and macrostuctures of these resources. The gap in our repertoire of possible descriptions of structures of content currently lies in the description of the creative processes that yield these structures. With the availability of such description guidelines, the practical means will emerge that allow us to draw the relations between the creative processes and any kind of process that builds on this creativity explicitly. In this way multiple definitions of meaning can be permitted to co-exist within the same framework. By putting people back into these processes at a fundamental level, we can redress the current technology-driven imbalance.

3.1 Adaptive System Design

Recent years have brought a silent revolution in the informatics community. With the growing influence of open standards advocates and the ever growing demand for interconnected functionality delivered over the Internet, an array of standards, protocols and technical concepts have created a new paradigm for delivering ubiquitous access to information. The **GUTENBERG 2.0** network will rely heavily on this set of now commonly adopted technologies which are associated with the SOA approach to developing information networks. Although this array of technologies is beautifully mastered by software engineers, helped by an ever increasing set of development tools, users are still left out by the sheer complexity of toolsets. As such, this timeframe looks somewhat similar to the mid-nineties when a communication officer still depended on technicians to publish corporate information on the internet.

The **GUTENBERG 2.0** network will strive to make the same difference to end-users as for the early adopters of Content Management Systems. End-users will be able to create information functionality by combining and authoring different types of services in ‘composite functionality’, which will reflect a particular workflow or way of sharing information. These composites are combined and interconnected to create a knowledge infrastructure. In this way, we can address one of the most important issues for achieving mainstream adaptivity: that such processes are available when and where they are needed, and by the appropriate person in the information provision chain. For both the public and private sectors, this represents a significant breakthrough and has the potential to overhaul adaptivity within emerging environments.

The approach we are describing can be called *adaptivity from scratch*. By building on recognised Design For All methodologies, systems should be built in such a way that the mainstream solution should be easily adaptable and extensible to add functionality for niche markets. As a result of the comprehensive lack of understanding of this concept at the fundamental design level, and strict deadlines to complete software projects, most adaptive solutions are ill-conceived and unlikely to meet the needs of the end-user. This often raises the question (though rarely explicitly) of whether the specialised needs of the niche market merit the effort involved in providing an adaptive solution.

Requirements never stay the same over time: requirements change for all users of any service. The end-user’s sight or other senses might deteriorate over time, their needs being met with appropriate features in accessible media. The differentiation of user requirements in general might grow, forcing the system to deal with a broader variety of processing possibilities with which it cannot cope. The processing system itself might in due time signal changes in memory requirements. The consumer base might be expanded to cater not only for visually impaired users, but also for dyslexic users. How can we anticipate fundamental changes like this? On the other hand, there exists a dynamic group of adaptive information producers who are pressed to keep up with the new media technology possibilities. The changing nature of requirements- and with that the potential design of any system- is a fundamental issue in the design of an inclusive world.

3.2 Adaptive Content Processing

It is possible to identify key trends in adaptive content processing that are likely to be of some importance in the coming years. There are two crucial principles that guide this work and the network partners have been carefully chosen to ensure that these principles are embedded at every stage of the research trajectory. The first principle is the clear need for *adaptivity on demand*. There are many different motivations for wanting to create adaptive content: be it legislative requirements [2], good practice [3], conformity with national guidelines [4], commercial imperatives [5] etc. In a sense the motivation in itself is a secondary consideration: what is required is a suitably flexible infrastructure to enable on-demand services to thrive. The goal of adaptivity on demand is to research fast and efficient services that allow converting inaccessible content into accessible and adaptable content: both for people having a need for accessible content (e.g. people with disabilities) and those who are under pressure to provide accessible content (e.g. publishers, public bodies). The **GUTENBERG 2.0** network will undertake research to put this requirement within the system architecture by analysing and implementing prototypes of a highly flexible middleware layer offering component based services based on coherent SOA architectures [6].

The second requirement is for adaptivity to be embedded within mainstream content creation and production processes at the earliest stages; that is, *adaptivity from scratch* [7]. Considering the move from accessible content processing to adaptive content processing can capture this principle, which goes beyond access on demand and should provide the highest level of both accessibility/usability and efficiency. In order to build extensibility into a system, the architecture should be such that every element used for processing the information is adaptable. This can be achieved by creating a representation layer which builds an object oriented structure from the information and which is free to adapt the meta-relationships and hierarchies intrinsic in that data genus. This is defined by identifying the parameters upon which the structure is built, and ensuring they are interconnected in such a way that promotes future adaptability without degrading the system: which is to say, using the right parameters for adaptive content processing. This should allow an as efficient as possible service provision for people with disabilities and other groups in specific situations which ask for alternative access (e.g. driving in a car, being in a busy train station,…) in terms of reducing the adaptation work for individual user groups by special service providers.

Most adaptive technology allows the user to access a document, and to read it following the "outer" structure of the original. But information also has an "inner" structure that allows the adaptive device to distinguish between a phrase and a measure, between a paragraph and a sentence, highlighting particular annotations, then the level of accessibility (and therefore usability) of the whole document will be greatly enhanced, allowing the user to move through it in the same way as those without impairments do when looking at a printed document, and following the same integral logic.

In an ideal world, all documents available in electronic formats should contain that internal structure using a standardised mark-up that benefits everyone. Highly-structured documents are becoming more and more popular due to reasons that very seldom pertain to making them accessible to people with disabilities. The move to XML related formats and associated standards for metadata have provided an impetus for far greater document structuring than before. Whatever the reasons behind those decisions are, the use of highly-structured information is of great benefit to a) different end-users accessing these documents b) with different devices, c) in differing locations d) at different times, e) in a different mood and so forth.

In recent years, the market for accessibility and assistive technologies has started to gain recognition in accordance to the raising awareness for the ICT potential and the need for accessibility on demand. The convergence of interests in "deep access" to content both in the accessibility field but also for enhanced accessibility for mobile users asks for ongoing research in adaptivity from scratch. The content provision and publishing sector has a need for "multi channel publishing" what means that the content has to maintain the mentioned inner structure for a fast and efficient reaction on market demands for content in varying formats.

It is clear that the integration of accessibility notions into mainstream technologies would provide previously unavailable opportunities in the provision of accessible multimedia information systems. It would open up modern information services and provide them to all types and levels of users, in both the software and the hardware domain. Additionally, new consumption and production devices and environments can be addressed from such platforms and this would provide very useful information provision opportunities indeed, such as information on mobile devices with additional speech assistance.

3.3 Adaptive Interface Design

Developing and maintaining adaptive content opens a broad variety of access to the same content. Adaptive Interface Design discusses new interface and therefore access opportunities based on the use of new devices in differing situations. Adaptive Interface Design in the age of the desktop focused on increased usability of a stable interface set-up. Accessibility for people with disabilities not being able to use the standard interface has been and obvious challenge for example. Content, although the idea of deep access is already around, tended to be designed for stable interfaces. Accessibility on demand, as described above, took over the task of giving access to non-standard interfaces. When moving away and beyond the desktop with new end-user devices into varying situations a stable interface non standard demands for access become standard asking for adaptive content processing to enable Adaptive Interface Design for deep access to content.

Modeling information can be separated into three phases; information retrieval, information representation and information reproduction. Retrieval concerns the perception of the information: once perceived, this perception is represented in some manner and can then be reproduced for the consumer. This continuous loop is the same for any producer/consumer relationship, where all consumers are also producers and vice versa.

Different users of the same content necessarily have different perspectives on that content. For example, to academics a book (even a work of fiction) is a reference source for their field. To a layman, reading a book is a leisure activity. To an author, the same book represents a means to communicate concepts. To a publisher, this versatile object is a unit of production in a wider supply chain. Given these multiple perspectives on something as familiar as a book, it is clear that one person's output medium is another's input medium.

4. Initial projects

This section describes an initial project for each of the key areas outlined in section 3 above.

4.1 Adaptive architectures for intuitive system design

How do we build systems that will help us survive in the digital age? Human survival is also required in the digital age. How should we imagine and decide which query is vital? Which transaction should not be missed? And how do we preserve these insights for posterity? How can we predict changing requirements and how do we make sure that anyone can participate in the definition of these changing requirements? Changing requirements not only imply a change in the definition of existing and known requirements. New situations, such as changes in environments – from economical to environmental - also need to be addressed, as do changes in personal and physical requirements. Everyone will get older and everyone can be unfortunate enough to lose certain perceptual abilities. How do we define the scope of the anticipation of changes we are trying to address? How could we model scope in such a way that it becomes a feature instead of a topic that is too difficult to tackle?

Artists communicate their thoughts through art and embody imagination. In the case of a live performance for example, various musicians communicate on the fly. It is the quality of the communication between musicians that is absorbed by the audience and decides if they feel inspired, respected and served. This is equally true for any content creators. As technologists, to ensure convincing system design it is important to consider the scope and the focus that we take on this content. The ability to see one another's requirements without too much destructive filtering is called interpersonal perspective taking. How then do we represent interpersonal perspective taking within a manageable framework that not only incorporates diverse wishes, but also locates these wishes within an organisational structure?

This may be possible by introducing harmonisation through complementarity. If we succeed in building a representation model that is based on communication instead of being solely based on stating individual facts, we would be able to create a model able to communicate with newly added content. Content that wasn't conceived and integrated in the framework beforehand. This content could bring its own model of meaning. But through the ability of the knowledge representation framework to communicate, the newly added content and its associated model of meaning could communicate with the models already in the content repository.

Through communication the new content and the existing content can express their deeper meanings. Through this communication a possible consequence for the newly created network of knowledge could involve learning. It could be possible that the deeper meaning will surface because of the possibility of communication between the various content entities and their models. These deeper meanings will remain submerged in a vast pool of models, meanings, convictions and so on for every aspect that we can think of, if there is no way of comparing these models. The only future for such a scenario would be additions of even more models, meanings and convictions. To allow comparison to occur, the absolutely minimal requirement is communication of features.

A framework that allows association of content, semantics of content and models that describe the ways in which any kind of user accesses content, has to allow harmonisation of the models that describe the content, the models that describe the semantics of the content and the models that capture the user's preferences, requirements and consumption and production behaviours. It is the interfacing between the user models on one hand and the content and content semantics models on the other hand, that determine the level of flexibility and freedom these users have in choosing themselves how they want to explore the content. It is the user who decides the aim of the exploration; it can be for pleasure, it can be learning. It can be various channels of interaction (horizontal relation) or levels of abstraction of interaction (vertical relation) with the content, the content semantics, the dynamics between both that influence or stimulate the user to choose it's aim for the exploration.

The aim is to conceive a universal language that introduces complementarity to achieve the harmonisation between the individual observer's observation requirements. Heterogeneous sets of observation requirements have to be allowed to exist in parallel, without inflicting damage each other and limiting the freedom of each others observation processes. Complementarity ensures that one particular observer can focus on one tiny aspect of the whole exposé, whilst the structural integrity that is required to be able to quickly gaze into another discipline's knowledge body remains intact.

Business opportunities arise from a demand that can evolve anywhere. Demand is a side product of a question which arises because someone or something in the audience or the serving side of the communication chain imagines the existence of a better or more suitable solution for a specific problem. The progress achieved in fulfilling this creative demand and turning the imagined solutions into useful products is what we would normally call innovation. Sustaining and preserving innovation in cultural manifests that show themselves as books, performances, applications, networks, communities, social networks and so forth could be considered a growing, evolving civilisation. Integrating these pieces of imagination into models and associating these with models of the individual beholder's requirements requires systematic and strategic attention to the actual stake-holder requirements, as expressed in educational, research and commercial terms.

4.2 Modeling granular addressability

The main goal is to research and develop techniques for generating small narrative units for inclusion in multi-platform environments with a view to enhance meaningful interaction with content and stimulate new forms of co-operative media experiences. This will research and develop new technologies to support this new kind of creativity, including both the content it creates and the environments it brings into being. At its heart is what is becoming known as Intelligent Content with its convergence of multimedia, web and knowledge engineering. Sometimes described as 'content that listens,' intelligent content is designed for the interactive age, reaching out from linear stories to possibilities that are as vast as the web itself. This kind of content is very new, and poses challenges for authors, producers, delivery systems and business models. Since re-use (re-mixing) of existing content is a feature of some work, it also poses legal questions for a media profession whose contracts have been largely drawn on a national and per-use basis. While existing projects focus on the professional creators, this work is also aimed at users, helping empower them to create and interact with media that are highly flexible and open-ended. Why should one want to address parts of content? One important use of this is when communicating information to someone else, possible with a comment added to it : "Look at this remark!", "What would the author mean with this?" etc. On printed material it is very common to underline or to highlight texts: on the web this cannot be done. There are some first attempts like www.fleck.com, which use a Flash layer for this, but this is far from standard and uses proprietary techniques.

Other applications which would become feasible when a more advanced addressing scheme were in place, would be for accessibility: DAISY has to identify parts (like sentences) by wrapping containers (spans) around elements in the source and assigning these an ID: this is needed to highlight text parts in sync with audio being played. If addressing in standard webpages would be made easier, it would become possible to add (several) accessible layers to these pages. In such a layer synthetic (or human narrated) audio could be added to the text and alternative text descriptions could be added to images.

4.3 Context-Specific Semantic Resolution and Personalisation

This project highlights the foundational research issues relating to the provision of ambient intelligent systems and services and suggests a research agenda on user-defined personal context representation and profiling management to underpin virtualisation and personalisation paradigms as a pre-requisite to the realisation of the vision of truly of assistive ambient and autonomic devices facilitating a range of secure pervasive- semantic-cooperative intimate systems and services.

It is widely acknowledged that information overflow on the web is a real problem; this has led to various approaches to focusing the provision of information and services to better match the users' needs, but to-date service selection and matchmaking based on the underlying tacit criteria acted upon by humans during choice-making is only weakly reflected.

Web-based interactive and collaborative social spaces (e.g. online information retrieval, online shopping, Wikis, Blogs, collaborative filtering/voting for issues/products/services, etc.) as commonly used by the current and emergent World Wide Web generations known as Web 2.0/3.0 or Social Web, will benefit from user-viewpoint mapping of the user's tacit context boundaries. This has motivated much of the research work [8] and others in the related sub-fields of Advanced Personalisation and Preference Engineering Technologies, including closely associated technologies of Usability Mining and Dynamic Usability Modelling to serve the co-design of Advanced Adaptive Interactive Systems, referred to simply as Intimate Systems [9].

Personalisation of information retrieval results and responses (e.g. from Embedded Conversational Agents during scenarios such as avatars offering advocacy, or, shopping assistance on the web) involving dialogue or multimedia presentation flow management have often tended to use, at best, simple model-based context recognition to personalise their responses to the user.

Advanced context-aware personalisation requires a deep-knowledge-driven sense of what in every individual user's mind might be the distinguishing factor(s) delineating his/her various specific contexts of interactions online. Dynamic updating of a user's preferences relating to each of their own implicitly observed personal contexts of interaction with the system is a key pre-requisite for offering truly context-aware systems and services.

Yet for most current personalisation systems this crucial task of updating user preferences amounts to little more than a simple additive process as in appending a list, e.g. with each purchase, to fit a priori ontologically-imposed models rather than building a dynamic preferences model, empirically–derived, based on longer term behaviour history and thus the experientially-deduced latent semantics and situated logic appertaining to each episode of a user's choice-making actions online.

Thus the rationale for our research agenda including systems for discovering the contextual cues that delineate the context boundaries which are implicitly distinguished in the minds of users, each according to their own private belief systems i.e. their own world-hood which is their own personal ways of seeing, patterns-of-situated-action, and, relating-to-context from which individual idiosyncratic patterns of preferences emerge.

5. Industrial context

European publishers are global leaders in their field, 6 out of the 10 biggest world publishers are European [10]:

- Reed Elsevier is Anglo-Dutch and leader in the professional fields,
- Pearson is British and leader in educational publishing,
- Bertelsmann is German and leader in trade publishing

The growth rate of the publishing market is quite stable but modest: in its "Media Outlook" study, PriceWaterhouse Coopers predicted the years 2004-2008 would generate a compound annual growth rate (CAGR) of 2% for consumer book publishing and 2.5% for educational and professional books including training in the EMEA area. However, this market has a strong business potential, whether with the prospect of the emergence of e-readers that allow the development of electronic books, or in terms of international trade exchanges, notably with emerging economies. (For instance, in 2006, Korea ranked before English speaking countries for the sales of rights of books of French publishers. Source: International Statistics of the French Publishers Association, 2007). Furthermore, books are obviously the fundamental vehicle of European culture, knowledge and languages, which the European Union (EU) seeks to promote [11]. The sector of books in Europe currently earns a turnover of 22 billion € [12] (probably 40 billion € expressed at the retail selling price), which means that publishing is the leading cultural industry in Europe. By way of comparison, the turnover for retail sales of films in Europe amounted to 11 billion € in 2004 [13].

Traditionally, the economics of information delivery has been subject to a trade-off between "richness" and "reach". Reach is understood to mean the number of people receiving the information, while richness refers to the amount of information communicated, the extent to which it can be customised, and the degree of interactivity between the sender of the information and the person receiving it. The implication of "unbundling" information from its carrier is that this traditional trade-off no longer applies. It is rapidly becoming possible to deliver large amounts of information to individual consumers in a highly targeted way and for a relatively low cost. Trends and developments in the international publishing industry have for some time been responsive to this reality.

However the sector is confronted with the media integration revolution that is based on the ICT revolution. This revolution provides many challenges of reusing, integrating and thereby multi channel distribution of content. It is not only to preserve the traditional book market; it is the challenge of using emerging opportunities of a changing culture of information usage which asks for according research and skill development that the sector can fulfill its key cultural role in the future.

6. Conclusion

Although the importance of publishing is recognised there are only very few comprehensive research attempts which intend to support this challenging process of change from a very stable, single medium oriented business to a multi media, multi modal, user driven and networked ICT industry. **GUTENBERG 2.0** wants to initiate research activities to accompany and support this process in its technical, economical and social dimensions.

7. References

[1] Hess, T. et. al.: Verbreitung, Anwendungsfelder und Wirtschaftlichkeit von XML in Print-Verlagen (2004) online at: http://www.wim.bwl.uni-muenchen.de/download_free/sonstiges/mreport_2004_02.pdf

[2] European Union, Council of the European Union: Accessibility of Public Websites -- Accessibility for People with Disabilities: Council Resolution, 2002

[3] Miesenberger, K.: Best Practice in Design for All, in: Stephanidis (ed.): The Universal Access Handbook, Lawrence Erlbaum Associates Inc. (LEA), 2008 (accepted; expected publication September 08)

[4] e.g. Behindertengleichstellungsgesetz: Regierungsvorlage (government bill): http://www.parlament.gv.at/portal/page?_pageid=908,848700&_dad=portal&_schema=PORTAL, 2007

[5] Darzentas, J., Miesenberger, K.: Design for All in Information Technology: a Universal Concern (Keynote), in: Andersen, K., V., Debenham, J., Wagner, R. (eds.): Database and Expert Systems Applications, 16th Internaitonal Conference, DEXA 2005, Copenhagen, Denmark, August 2005, Proceedings, Springer LCNS 3588, Berlin/Heidelberg 2005, pp. 406 – 420

[6] MacKenzie, C.M., et al. (2006) Reference Model for Service Oriented Architecture. [cited 2008 01-16]; Retrieved from: http://www.oasis-open.org/committees/download.php/16587/wd-soa-rm- cd1ED.pdf

[7] Crombie, D., Lenoir, R., & McKenzie, N., (2003) Accessibility from scratch: how an open focus contributes to inclusive design, Amsterdam

[8] e.g. as in MobiPETS-GRID, Badii, Hoffmann, Heider 2006

[9] e.g. Badii et al 1995-2008

[10] Source: Ranking by Livres Hebdo on the occasion of the 2007 Frankfurt Book Fair: http://www.livreshebdo.fr/cache/upload/pdf/FrancAng.pdf

[11] Particularly through the creation of a new portfolio under the responsibility of Commissioner Leonard Orban

[12] Figures for 2004 of the FEP: http://www.fep-fee.eu

[13] IVF (International Video Federation): http://www.ivf-videon.org/site/share/dlm/Files/European_Overview_05.pdf

uWEAR: Wearable computing for you

Roselyne van der Heul, Hugo Verweij & Richard van Tol
Faculty Art, Media & Technology, Utrecht School of the Arts
Utrecht, Netherlands
{roselyne.vanderheul}{hugo.verweij}{richard.vantol}@kmt.hku.nl

Alexandru Stan, Michael Weber & George Ioannidis
IN2 Search Interfaces Development Ltd
Bremen, Germany
{as}{mw}{gi}@in-two.com

Abstract

The WearIT@Work project was set up by the European Commission as an Integrated Project to investigate "Wearable Computing" as a technology dealing with computer systems integrated in clothing. It is the largest project worldwide in wearable computing with 42 partners. WearIT@Work sets the stage for the applicability of wearable computer technology in various industrial environments. WearIT@Work novel computer systems support their users or groups of users in an unobtrusive way. This allows them to perform their primary task without distracting their attention and thus enabling computer applications in novel fields. One of the major goals is to investigate the user acceptance of wearables and the project established three take-up actions, one of which was the uWEAR project. This article describes the system developed by uWEAR, the results of the user tests and outlines the planned implementation.

1. Introduction

While other pilots of WearIT@Work [1] chose a more industry oriented approach to wearable computing and its advantages, uWEAR wanted to address users on a more personal level and investigated how new technologies could enrich their everyday life. By adapting and extending existing WearIT@Work wearable components, uWEAR developed navigational services for visually impaired users. The interfaces designed in uWEAR should allow the user to efficiently get the required information whenever necessary, while minimizing interference with current actions. uWEAR is not meant to replace current tools, such as white canes, but rather augment them in order to empower the users and give them more independence. uWEAR makes it possible to, for example, get route guidance to a place that was never visited before, or always find out the current position - two not so trivial tasks for the blind. Also, the system will be enhanced and customized for other target groups, for example cyclists, who require the same level of unobtrusiveness when it comes to navigation. The user tests for this phase of the project focused exclusively on its primary target group: visually impaired people. This article describes the system developed by uWEAR, the results of the user tests and outlines the planned implementation.

2. Objectives

Over the past few years, digital navigation technology has rapidly found its way into daily life through a wide range of commercial applications and tools. These do not only include popular car navigation systems but also applications in entertainment, tourism, education, health and games. Unfortunately the majority of these developments are inaccessible for the visually impaired, for whom way-finding in indoor and outdoor public spaces, especially those which are new, often provides many challenges such as:

- Navigation and orientation challenges: getting to a specific location can be difficult due to a lack of information about position, direction and location;
- Safety challenges: moving around safely is difficult because of obstacles and dangers on the route that cannot easily be noticed and/or avoided;

- Experiential challenges: due to the lack of visual stimuli important, interesting information about the surroundings is missing.

Previous projects have already shown that blind users can benefit from digital navigation technology, but also that there is still much room for improvement and refinement. Quite often the focus was on the technical implementation and getting the relatively new technology to work. The consequence was that there was too little focus on the user. uWEAR has tried to turn things around and to start as much as possible from the user-perspective.

uWEAR has examined a user-centred design approach for testing the application of wearable computing in outdoor and indoor navigation scenarios for visually impaired users. By adapting existing WearIT@Work components and combining these with other interfaces, it has extended the application of the WearIT@Work results to fit to users with special needs. The final outcome of uWEAR is a practical understanding of how wearable computing interfaces can be advanced in order to cope with the requirements posed by persons with special needs. Furthermore, specific adaptive user interfaces for visually impaired persons have been developed and tested for outdoor and indoor navigation scenarios.

3. The uWEAR System

3.1 Approach

The approach followed by uWEAR is adapted from User Sensitive Inclusive Design that bases its methodology on the Design for All / Universal Design movement. An iterative development cycle was chosen to ensure that a maximum of user requirements were incorporated and validated in the technical solution, thus improving acceptance of the developed technologies. Especially in the case where new computing paradigms such as wearable computing are to be developed, a development model with short iterative cycles is a prerequisite for including the user-perspective into the design of these technologies.

The uWEAR approach incorporated the usage of the given wearable computing components developed in the WearIT@Work project and the development of emulations or mock-ups for testing those with the target user groups. This preliminary concept of operation consisted of a set of practical tasks to perform in outdoor and indoor navigation scenarios, and aims in validating the user requirements.

In total three phases of user evaluations were carried out: (a) concept user test phase, (b) pilot user test phase and (c) prototype user test phase. Similar to that, also three phases of developing and refining the user interfaces were carried out: (a) concept development phase, (b) pilot development phase, and (c) prototype development phase. The different phases interacted together to deliver insights in user needs and requirements that functioned as the basis for the next phase in the design process.

The requirements and design of the system were adjusted continuously at the beginning of each of the three phases described above. Major adjustments came from the hardware side. The original design involved the usage of a wrist worn wearable computer (Zypad) and a headset. After the first development iteration the refined requirements obliged these components to be replaced. Instead of the Zypad the OQO was used, while the headset was replaced by a wearable speaker and clip microphone.

3.2 Components

The uWEAR system uses several components from within the Open Wearable Computing Framework (OWCF). Additionally, these were adjusted and integrated with other components off the shelve (COTS) or specially developed ones. A brief description of the different components is listed below.

The TZI SCIPIO Winspect Glove [2,3,4]: a general-purpose wearable input device that was developed by the WearLab at TZI Bremen. The physical appearance is of a lightweight fingerless glove with two buttons that are fitted around the finger. The glove was designed in such a way that it does not interfere with the way a user performs regular tasks. It is equipped with a small microcontroller, an acceleration sensor, an 125kHz RFID reader, Bluetooth module and battery pack (up to 8 hours of usage time). In order to integrate it with the uWEAR system, the device drivers included in the WUI-Toolkit (described below) are used. Before usage, the glove needs to be calibrated. Through custom micro-gestures and the finger buttons, the glove can be used to browse menus and make selections.

Figure 1: TZI SCIPIO Glove

The uWEAR team developed a wearable speaker. We tested different speaker designs and decided to use a small and cheap off the shelf available speaker that can be connected trough a mini-jack connection and that works on its own AAA-batteries. We integrated this speaker into a small reflective, waterproof and magnetic cover that can be easily and flexibly attached to different pieces of clothing.

The computing horsepower is provided by the OQO e2 wearable computer, weighing just under 500g and measuring 142mm(W) x 84mm(H) x 26mm(D). The OQO features a 1.5GHz processor, 1 GB of RAM memory, allowing it to run Windows XP with no problems. Connectivity is assured via Wireless LAN and Bluetooth. The device also provides an USB port and an i/o audio jack. For the outdoor scenario the current coordinates can be provided to the system using any common GPS receiver. For our user tests the GPSLim236 from Holux was used. Bluetooth is used for receiving its GPS-data.

A major part of the output interface of the system is the audio menu, which provides sound-only feedback of the system and which is controlled by the data glove. The looping menu is designed so that it supports not only novice users, but also more experienced users and even expert ones. For this, each menu item consists of three parts. The first part, a generic beep which pitch and timbre is based on the number of items and depth in the menu, allows expert users who are already familiar with the menus to browse solely on pitch. The second part consists of a short line of speech, announcing the menu title. The third part consists of a more lengthy description of the menu item, describing what the menu item is used for. This allows novice users to learn the function of each menu item. Each part is followed by a small silence.

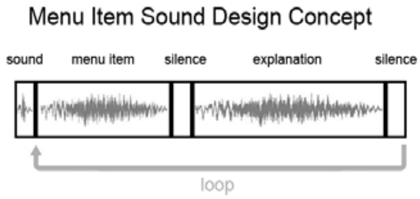


Figure 2: Audiomenu Item

As an alternative input interface the, the system uses the speech recognition component of the OWCF developed by Multitel. This allows users to interact with the system by issuing simple commands such as "previous", "next" and "select" or make selections according to the menu item names. Those are available within a grammar, which is automatically created and updated for the current menu screen.

The WUI-Toolkit [5] is a framework to support and ease the development of wearable user interfaces (WUIs). The toolkit presents a first step towards a model-driven UI design approach in wearable computing that allows even non-UI experts the generation of usable and context-aware WUIs. Based on an abstract model of an envisioned user interface that is independent of any concrete representation, the toolkit is able to generate a device- and context specific UI for a given wearable computing system at runtime. The toolkit features the ability to use available context sources and can automatically adapt generated interfaces to maintain their usability.

As the outdoor navigational engine, the Maptrip software from Infoware was used. Maptrip provides an SDK that allows using basic services, such as providing a route description between a specified starting position and a destination.

For the indoor scenario the uWEAR team built the navigational software. It provides instructions based on the current coordinates and a map of the building. For this scenario the floor of an indoor space is constructed of multiple RFID floor tiles – with each tile having four RFID tags. Each RFID tag corresponds to a certain coordinate on the map. The system estimates the current position of the user based on the tags read in a short time interval and the user's previous position. In the user tests, the current coordinates were provided using mock-ups of the RFID tiles and RFID reader in the cane.

3.3 System architecture and design

As mentioned before, uWEAR was built in a modular fashion such that various components can be replaced easily as upgrades become available. The figure below gives a representation of the high-level system architecture:

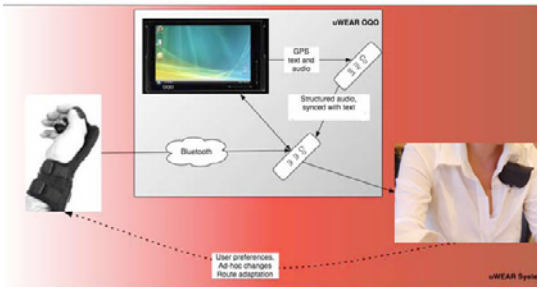


Figure 3: Architecture

The user can navigate the audio menu using the Dataglove or the Voicemouse. The audio menu is built using the WUI-Toolkit. The user can then select a destination. The navigational software computes a route to this location based on the map and current position (GPS coordinates in the case of the outdoor scenario). Audio instructions are then prepared and given to the user via the wearable speaker. At any time the user can make route adjustments or make other changes to the system.

The WUI-Toolkit requires JVM 1.4 while the Maptrip software SDK is using C++. Between the two a specially designed socket interface is used. The rest of the application is written in Java 1.6.

4. Testing of prototype

As explained above, a total of three tests were carried out with 3 to 6 users each. The users were both male and female, between 18 and 48 years old, having different experience level with computers. Before each test, each user received a brief personal training into the working and use of the different components. The actual tests consisted of a number of tasks that the user had to fulfil, varying from simply setting a route to a specific location, to actual way-finding outdoors and indoors on the basis of instruction provided by the uWEAR system. After the tasks had been performed by the test-user, an evaluative interview took place in which user feedback on the system as a whole and on each of the different components was collected.

The first user test focused on the usability of the concept menu structure and the general user-demands regarding navigation scenarios .The results of the first user test were fed into the design of the pilot system including the audio menu, input devices, GPS software and hardware and a pilot-version of the wearable speaker. This integrated system was tested in the second user test, leading to recommendations for further refinements per component and for an optimal integration of the different components. The testing of the final prototype lead to further insights and recommendations for future development of a marketable integrated system.

The user tests were conducted in cooperation with Bartimeus. Bartimeus is the largest organisation for visually impaired people in the Netherlands with 2000 employees, 12000 clients and 16 different locations. The uWEAR user tests were conducted in and around the Bartimeus office in Utrecht by employees of HKU and IN2 who received advice and support from employees of Bartimeus.

The last user test evaluated the final prototype of the system that took into consideration the recommendations of the previous tests. In this context the Audiomenu, the Dataglove input interface, the shoulder speaker, outdoor and indoor scenario were tested. Most users were happy with the improved responsiveness of the Audiomenu to the gestures or speech input, such as the responsiveness of the Audiomenu to the gestures and speech input, or the new zero-position of the Dataglove. They also provided some more suggestions, such as having an "on/off" button for the Dataglove. Additionally, the new testers noted that, as observed in the previous tests, the innovative structure of the looping menu is very intuitive and easy to navigate.

The selection of a shoulder speaker was a very important part of the user test. Several models were presented and the users had to choose the one they liked most. In the end the selection that proved to be most popular was the magnetic-speaker. The users found it very pleasant that one can adjust the mounting position very easily and that it is also very discrete. As it turns out, fashion and general appearance does play a role to this user class. Most users would like to have the speaker connected wirelessly, using Bluetooth or a similar technology.

Testing of the outdoor scenario showed the need of using navigational software that is designed for pedestrian use. Nevertheless, the test also revealed that the system does indeed blend very well with the existing navigational tools of the visually impaired people: the system gives the instruction 'first street on the right', the user gives this instruction to the dog and the dog selects the first street on the right. Another participant that was using a white cane noted that while the instructions from the system sometimes came quite early or too late ('turn left') the participant could decide for herself whether or not it was really the right position to turn right by combining the audio instruction with echo-location techniques.

Testing of the indoor scenario was based on a mock-up of an actual RFID tile room. For logistic reasons it was not possible to have real RFID tiles and readers for this user test. As the users were walking across the room the coordinates were manually entered to the system. The test was concerned with the way to provide instructions to the users and the type of instructions that would be optimal for them. The user feedback was very positive regarding such a system. The users enjoyed the extra information that they received apart from the strictly navigational instructions ("kitchen on the right").

5. Initial conclusions

User Centred Design is an effective approach to develop this kind of wearable applications for user groups with special needs. The different stages in the development and user test cycles allowed us to include many valuable recommendations and requirements from the user-perspective.

In general it can be stated that the WearIT@Work components selected for the uWEAR project can be integrated effectively into a system that provides valuable navigational services to the special user group of Visually Impaired People.

The user test made it very clear how already existing technologies (echo-location, cane, dog and technology) can be effectively integrated with 'newly innovated' technologies (Dataglove, wearable computer, wearable speaker, Voicemouse, audio-menu) into one new system that is easily accepted by the user while really providing an added value. By avoiding replacement of trusted technologies but allowing for further improvements by adding-on to these, a new system can offer an optimal service.

Our initial assumption on the importance of a good audio-based menu for the success of a uWEAR type of technology, was confirmed in the tests. The importance of paying sufficient attention to the audio-design applies especially to this specific user group of Visually Impaired People, but it is likely to also apply to a wide range of other technologies that include sound as an important component of their interfaces.

The test-users also highly appreciated the attention given to comfort and fashionability. The test confirmed that wearable technology should be as comfortable, flexible and unobtrusive as possible. Although our target group is visually impaired people, the look and feel of the system remains important. Privacy is another important issue to further integrate into a final design: some users were concerned that others in the vicinity during the route set-up phase could listen into potentially sensitive information, such as home address and future destination.

6. Future work

Based on the results of the last user test further improvements can be made to the system. Firstly, Maptrip should be replaced with a better COTS navigational engine SDK that is designed for pedestrians. This should cater for the complaints received regarding the usability of the system in the outdoor scenario. Having the system build in a modular fashion allows for such upgrades to be implemented without much effort.

In the indoor scenario, the algorithm that computes the current position based on several RFID tag reads will have to be implemented. The algorithm uses an overlay of the RFID tags based coordinate system, in order to compute the position given certain probabilities of a successful or unsuccessful read and the previously determined values. Tests with actual hardware will then evaluate the prototype.

Since users reported that they would like to be able to use their everyday devices for such a service as the one provided by uWear, it is worth searching for possible optimisations and for ways to relax the system requirements. This would lead to the replacement of the OQO and the current software with software that can run on conventional mobile phones. This would make it easier to use the uWEAR service for most people and it would also mean a significant cost-reduction.

Further possible improvements would be giving users the option to receive also the navigational instructions (instead of only the audio-menu instructions) in the form of spoken instructions or bleeps. This would increase the speed of information provision and retrieval and increase the level of privacy for the user

Furthermore, one can investigate the possibilities of tactile feedback, i.e. connecting buzzing/vibrating components that users can wear (for instance around their wrist or on their shoulder) which provides navigational instructions in the form of vibrations felt by the user. This could further increase the unobtrusiveness, privacy and speed of the instructions.

Acknowledgment:
This work has been funded by the European Commission through IST Project WearIT@Work:

Empowering the Mobile Worker with Wearable Computing (No. IP 004216-2004).

7. References

[1] IST Project WearIT@Work - Empowering the mobile worker with wearable Computing (No. IP004216-2004). http://www.wearitatwork.com, 2008

[2] Hendrik Witt, Rüdiger Leibrandt, Andreas Kemnade, Holger Kenn: SCIPIO: A Miniaturized Building Block for Wearable Interaction Devices. 3rd International Forum on Applied Wearable Computing (IFAWC), Bremen, Germany, March 15-16, 2005

[3] Michael Lawo, Hendrik Witt, Holger Kenn, Tom Nicolai, Rüdiger Leibrand, (2006), *A Glove for Seamless Computer Interaction - Understand the WINSPECT*. In Proceedings of the The Smart Glove Workshop, pp. 3-8

[4] Holger Kenn, Friedrich van Megen, Robert Sugar, (March 12 - 13, 2007), *A glove-based gesture interface for wearable computing Applications*. In Proceedings 4th IFAWC 2007, Tel Aviv/Israel.

[5] Hendrik Witt, Tom Nicolai, Holger Kenn, (29th June, 2007), *The WUI-Toolkit: A Model-driven UI Development Framework for Wearable User Interfaces*. 7th IEEE International Workshop on Smart Appliances and Wearable Computing (IWSAWC), Toronto, Canada.

2020 3D Media Panel on User Intimate Requirement Hierarchy Resolution for Work-flow Design and Implementation in (Post)-Production & Distribution

Organized by

Intelligent Media Systems and Services Research Centre

School of Systems Engineering, University of Reading (UK)

Coordinated by

Atta Badii, David Luigi Fuschi (University of Reading, UK)

AXMEDIS 2008: proceedings of the 4th International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution. Workshops, Panels and Industrial Applications. Florence, Italy 17-19 November 2008
edited by Paolo Nesi, Kia Ng, Jaime Delgado
ISBN: 978-88-8453-811-6 (online). ISBN 978-88-8453-810-9 (print). © Firenze University Press, 2008

User-Intimate Requirements Hierarchy Resolution Framework (UI-REF) in Work-flow Design for 3D Media Production & Distribution

Atta Badii, David Luigi Fuschi
Intelligent Media Systems and Services Research Centre
School of Systems Engineering
University of Reading
RG6 6A,Y UK
{atta.badii, d.l.a.fuschi}@reading.ac.uk

Abstract

This paper presents the User-Intimate Requirements Hierarchy Resolution Framework (UI-REF) based on earlier work [13-16, 95-104] to optimise the requirements engineering process particularly to support user-intimate interactive systems co-design. The stages of the UI-REF framework for requirements resolution-and-prioritisation are described. The UI-REF framework is then applied to the case study which is focused on process modelling for workflow design and implementation for 3D media (post)-production and distribution. UI-REF has been established to ensure that the most-deeply-valued needs of the majority of stakeholders are elicited and ranked, and the root rationale for requirements evolution is trace-able and contextualised so as to help resolve stakeholder conflicts. UI-REF supports the dynamically evolving requirements of the users in the context of digital economy as underpinned by online service provisioning. Requirements prioritisation in UI-REF is fully resolved while a promotion path for lower priority requirements is delineated so as to ensure that as the requirements evolve so will their resolution and prioritisation.

1. Introduction

In any process of system design and development, the requirements engineering phase is crucial as its deficiency will often result in a high probability of overall failure of the resulting system to meet the priority needs of the users to a satisfactory level. This principle has been established for a long time and is widely reflected in literature and standards, particularly for software development [18, 19, 21, 27]. Despite this, the key role of the requirements engineering phase is often under-estimated and poorly executed. In some cases the requirement engineering processes that are followed appear to be overly formalised but substantially failing to be *"user-intimate"* by which we mean a user-centric process closely integrated with the testability framework and the formative usability evaluation [16, 15]. UI-REF is rooted in earlier work on user-intimate systems [13] and usability evaluation [17] along with the most relevant work in the requirements engineering and project management domain [1-11, 34-82].

In this paper we set out a description of the essential features of UI-REF as already documented and applied extensively to several projects [18, 28, 29, 31-34] and supplemented by well-established techniques such as nested video-ing of user interviews and scenario walk-throughs [20, 15, 28, 31] etc. which represent our routinised approach to User Requirements Engineering, and more importantly, to requirements prioritisation, and, stakeholder conflicts resolution. We say routinised as even although requirements should be *"elicited"*, most often they have to be systematically extracted and inventoried; starting form an extensive data collection activity involving all possible stakeholders, literature, similar systems, feasibility studies, market analysis, business plans, competing products / services / systems, research state-of the-art and overall domain(s) knowledge [26].

2. Ensuring shared understanding

In the context of system design, when it comes to the representation of stakeholders' needs, in a format suitable for the development team to have a clear and unambiguous reference, really shared meaning across the participants is a pre-requisite to the integrity of the process end-to-end. It is at this starting point that the requirement engineering methods can help or hinder the capturing of the ground truth about what really are the most-deeply-valued needs of users; thus avoiding various forms of mismatch that may otherwise arise way downstream causing untold amounts of avoidable wasted resources and users' disappointment [1]. To overcome communication barriers due to various sub-languages, sub-cultures, prejudices and articulation-theoretic difficulties that might well be encountered in a given stakeholder community, it has proven essential to start any discussion that will lead to either requirements elicitation or any other system development related activity, by defining the common ground for understanding via the definition of a commonly accepted glossary of terms [21]. This initial step has often proven so crucial that failing to find such common ground could lead to unsatisfactory results for all concerned. A more enterprising strategy that would go beyond this is the UI-REF with Virtual Users approach [15, 95, 101] in which UI-REF is empowered with an online simulation environment, possibly supported by a (mini)CAVE or video-gaming as appropriate, to allow direct user experience with the intended system which is being thus co-designed, through user's live usability responses to its features as the user interacts with them during process enactment scenarios. Nested (video) recording of such online user feedback provides a valuable source of formative evaluation as well as allowing a real-life experience-based confirmation loop re the user's shared meanings and preferences as both the user and the target system are embedded within the simulated prototypical environment of the target application domain. If resources permit it and should the complexity, dynamic interactivity/bio-feedback and safety considerations (as in the aerospace sector) which may be required by target application domain necessitate this level of investment of effort, then this essentially deeper, simulation-empowered, version of the UI-REF would be a good candidate as a user-centric co-design requirements engineering approach as it also naturally supports the dynamic systems development methodologies such as the agile evolutionary approach.

3. Key stages in UI-REF

As anticipated UI-REF addresses the requirements hierarchy elicitation, analysis, generalisation and resolution of users' needs priorities taking into account that what is more "intimate" to a user is also what is perceived as more relevant in the user's own context. Thus once preliminary steps have been completed, it will be necessary to analyse the domain, context and objectives of the requirements collection and analysis.

To this end it will be necessary at least to: ***i)*** agree and set out the list of domain prototypical entities or *actors* and *objects* as stakeholders; ***ii)*** define the characteristics of the prototypical entities which will have to include any entity that are involved, i.e. implicated, in any way whatsoever in the typical usage arena envisaged for the target system (i.e. need to set out respective *capabilities*, *roles*, *rights*, *responsibilities* for each entity i.e. each actor/object/process etc); ***iii)*** establish the generalisation ontology of usage contexts each related to their distinct Man-Machine interfacing features as needed by the relevant user (sub)groups in their target prototypical scenarios; ***iv)*** define the key differentiators of usage contexts (context switches) and the prototypical actors' needs hierarchies in each of the identified prototypical target context-scenarios; ***v)*** define the prototypical workflows and state diagrams, thus establish the domain (sub)-goal and (sub)-task hierarchies; ***vi***) deduce the user's needs priorities in terms of ICT-enabled features to facilitate user's task fulfilment in each situated context-scenario of the application domain as identified and demarcated (situated-usage-class) under respectively **iii** and **iv** above.

Such steps can also be grouped in terms of the kind of understanding and knowledge that they provide about the contextualised system to be designed. Besides, they can be more finely-grained and profitably grouped to provide a valuable level of abstraction and management of detail which is very important in deep-inspection / introspection requirements elicitation. This aspect is stressed as in our view no system can be fully described in isolation from the usage-context – when adopting formal description methods the context will be somehow represented by boundary conditions or some kind of constraints [13, 15, 28, 29, 31-33].

4. Managing requirements complexity and prioritisation evolution

Such user-intimate approaches often yield a vast amount of raw data and without appropriate abstraction and context layering, to reflect the natural partitions within the domain, one can end up with a forest of data but little actionable insight as to the most- deeply-valued needs for most users belonging to each of the target usage-context types within the spectrum of usage-context classes to be addressed by the target system. Specifically in identifying all domain objects and actors and delineating the boundaries of roles and responsibility spaces, rights/privileges for each actor and/or object in the domain, we are essentially negotiating a phenomenological analysis of the domain with the users. This process will also serve to expose articulation-theoretic, pragmatic, sub-cultural and sub-linguistic variations and their influences which are noted and used later in requirements analysis [20, 21, 22, 25-27].

Once this has been done, by specifying domain knowledge, taxonomies and a tentative ontology for the domain and negotiating this with various user communities, it is possible to conclude an appropriately partitioned ontology of the world of the users for whom the system is intended (i.e. use-context-language-of-things-patterns-values-user's preferences) [13, 15, 17, 19-21, 26, 27]. This is a crucial step for it at once serves to crystallise the relevant domain knowledge for the requirements engineering practitioner as well as complete a framework of reference descriptions for all who are implicated in the process [21]. Such generalisation ontology serves as a values expression language of the most-deeply-valued-needs for various usage-contexts; this is to be periodically negotiated and confirmed to keep up with the dynamically evolving values and thus changing needs of each class of users implicated in each use-context-type as they are to be the masters of their ICT servants. Such domain ontology is an ethno-methodological aid for deepening and widening mutual understanding, and, in the process revisiting and clarifying the deepest needs of each user, their strengths, weaknesses, what pains, frustrates, irritates and/or pleases them in accomplishing their (sub)goals and how they feel they could be best supported in accomplishing their tasks; and in fulfilling their priority needs including the need for self-determination of their interaction modalities with their servant ICT; their ways of seeing their world and their patterns-of-relating to people and things that they would have to interact with in accomplishing their routine goals in the target application domain. It is important at times to invite (or provoke) the users or their proxies to introspect aloud their feelings related to their needs and wants, to delve deeply into their own *"value-language"* to help them articulate their really essential needs; this is all about helping unpick the real from the perceived, imagined and illusory; helping the user distinguish the surface forms, the deep forms and the sub-texts in their ways of seeing values-affordances and patterns of relating to their own feelings about the things in their world-hood.

Building on this increasingly deeper understanding paves the way for formalising the domain knowledge structure including tacit knowledge, causal, processual and structural knowledge thus adequately specifying the domain knowledge structure. This comprises a most important element of the experientially-derived tactical/ strategic problem solving knowledge. The domain knowledge is clearly the provenance of the various user-classes, (as distinguished by their usage-contexts) who are to use the system in pursuit of their everyday practice. It is these *communities of practice* who are expected to make available to their requirements engineers and other stakeholders their domain knowledge so as to promote deeper understanding of their domain requirements including end-to-end interoperability and meta-operability across all implicated sectors. This domain knowledge captures the essential epistemology (structure, topology of knowledge of the application domain that is relevant to various situated contexts of usage). This includes the user's application of their working skills and knowledge in accomplishing their everyday target tasks that are the focus of the requirements engineering exercise.

As some of the requirements of the application domain may broadly contribute to usability but actually make little sense for a significant subset of the users depending on the pragmatics of their particular usage-context e.g. their daily (sub)-tasks or the way these are currently routinised within their legacy systems, it is clear that requirements need to be contextualised with the underlying pragmatics. This will index into a typology of the various environments that may be encountered in actual usage depending on the variations of semantics and syntactics of their legacy systems including their workflow, processes and services environment. This essentially captures the applicable semiotics overlay i.e. the subtle variations on the same theme of requirements that imply the need for provision of a particular degree of customisation re certain requirements so as to be able to potentially deliver different genres of the same system to best fit different user (sub)-sectors and variable organisational legacy systems.

In eliciting the domain knowledge we can establish the goal structures knowledge for the application domain. This has to include workflow and in particular distinguish the needed functionalities that relate to solo and/or team work and would facilitated any elements of planning, acting, execution monitoring, failure recognition, plan repair and recovery management that are interleaved by each user and integrated in teamwork within the application domain [5-11]. Thus this essentially captures the teleology of the actors within the domain to ensure support for global cooperativity facilitated by ICT-and-processes harmonisation i.e. end-to-end inter-meta-operability assurance [95].

The above layers of elicited and negotiated requirements knowledge would correspond to specific situated usability-sensefulness objectives for the target system to be best fit for each of its intended usage-context types. The above process essentially lays the common ground and foundational knowledge to enable effective elicitation of the requirements involving triangulation using a variety of methods, techniques, instruments and modalities ranging from conducting semi-structured interviews, work study and nested-video-assisted or otherwise (simulation)-augmented scenario-tasks-walk-through as well as various ethno-methodological approaches [20] to overcome users' problems in articulating their real needs, wants and preferences [21, 22].

The resulting body of elicited and negotiated requirements knowledge is subsequently analysed, validated and refined through various cross-consistency checks [21] and filtering stages [13 ,15] to ensure correct understanding of applicable pull (user / practitioner / market-initiated) and push (technology-initiated) factors influencing the requirements and the associated gaps arising from *State-of-the-Market* (SoM), *State-of-the-Art* (SoA), *State-of-the-Practice* (SoP) . This would allow the requirements engineer finally to refine, partition and prioritise the elicited requirements such that these will not entail either too ambitious a leap in the required innovation given the scope of the project, or worse still, a "re-invention of the wheel" i.e. failure to integrate "*Components Off The Shelf* (COTS)" as appropriate. The net result of tentatively prioritised requirements has to be referred back to the user-classes for confirmation, validation, refinement and re-partitioning as appropriate before being bundled as use cases with associated usage-contexts and test cases to form the basis of the specification, design, implementation and integration of the initial prototype. This would in due course need to be subjected formative *"living lab"* usability evaluations, re-engineering and refinements in the normal course of iterative, evolutionary co-design given the fact that it is certain that, despite all effort, still some requirements (especially for large systems) may not have been correctly understood / specified / prioritised in the first instance [1 ,19, 26], thus naturally leading to evolutionary revision and update of requirements and related specifications.

Everyday integration of users with an ICT system invariably creates a complex system irrespective of the scale of the task or system. The integrated human-machine system must possess high actability in the sense of maximally supporting each user's desired degrees of freedom in task execution and goal accomplishment with a spectrum of passive, reactive, proactive ICT functionalities at the disposal of the users to support their own life-style/work-style as they wish. The integrated system must also be highly sense-full in that it must support the teleological constructs of the domain, fit the established business process practice logic and lend itself to easy routinisation within everyday work patterns as preferred within the target environment. In short it must aspire to blend in, at the user's initiative, seamlessly and unobtrusively within the user's domain; be part of the solution not part of the problems or a solution looking for a non-existent problem; it must help users, to the extent that they like, to get done what they need to get done in the way they prefer to get it done.

By establishing the tasks and related goal structures for each agent, actor, entity, object, it will be possible to elicit the needed ICT functionalities as required to support the fulfilment of each task by each user. In due course after the list of all such functionalities are derived, compiled, filtered, partitioned and prioritised. The results will allow use-case, test-case pairs linked with each usage-context-type to be specified to inform the design specification of the framework architecture that is to deliver the target family of closely related usages i.e. the target usage-class-spectrum for the intended market segment which would allow some core ICT capabilities for the domain as a whole as well as some usage-context-specific support so as to allow seamless adaptation as the user's usage-contexts change within the target usage spectrum [95].

5. A quality oriented process

Given the growing complexity of ICT systems the quality process has to be integrated from an early stage. There are several methods suggested for this; both at the level of software engineering and project management [1-11]. In this respect UI-REF accommodates an auditing process aiming to maintaining requirement owner trace-ability and transparency of requirements.

The composition of the user community will have to be studied and if there are heterogeneous sub-sectors of the user community that can be distinguished on the basis of some prevailing usage environment and/or business/usage practice logic factors etc then the representative from each such distinct sub-groups of users have to be included in all processes of needs elicitation throughout the UI-REF implementation so as to ensure that the needs of all stakeholder groups are adequately considered.

The interview content and processes must be monitored to minimise prompting and bias by the interviewers whilst permitting appropriate dialogue to clarify any questions on demand introducing supplementary questions/visualisation/simulation if and when it is necessary to do so to seek clarification of a given response or to support user's introspection. Equally it is important to ensure that a hierarchical audit trail of the first-expression and subsequent modifications of each requirement and its evolution (root-question, birth-point, sponsor(s), modifications, priority-promotion / demotion, deletion) is maintained so that the rationale for the existence of each requirement element can be re-visited at any time. This is to ascertain which question led to the requirement (root-question), and the who, where, when, why of how each requirement was first expressed-justified (birth-point, parent-sponsor) and who was responsible for its evolution and how many sponsors/subscribers it has had in the user community etc. Such semantic mark-up of the *root-rationale* of each requirements element can be deployed within a multimedia database that could be designed for semantically-cued collaterally-and-multimodally indexable retrieval of any requirement element [13]. This should be able to support queries re spatio-temporal-and-media type evidence for the root rationale that has been recorded by all who have commented a given requirement historically and hierarchically[1] across space, time and modalities[2]. Thus the ontological, epistemological, semiotic and teleological domain knowledge extraction and partitioning as may be mediated through dedicated scenario exemplars has to be visited for each distinct user-type and linked usage-context-type(s) and the overall requirements knowledge is thus aggregated in a trace-able fashion. Equally all instruments to be used to elicit the requirements will have to be piloted to maximise the relevance and understanding of the questionnaire and responses in each case.

6. Requirements Priority Categories

UI-REF advocates that the requirements are classified into the following descending-order priority categories for implementation; these range from *mandatory*, as the highest priority class, through *desirable*, as the medium priority class, to *optional*, the lowest priority class [30, 31]. Mandatory requirements are expected to remain relatively stable. Migration from desirable and even optional categories into the mandatory ones can occur in the light of usability evaluations of the first prototype and market-technology updates that are expected typically midway through the lifecycle of the project. Prioritisation of requirements is deduced from careful analysis of user-stated priorities which can be aided also by a consideration of Purpose-Hurry-Frequency Criteria set re degrees of intimacy and immediacy of the required services and patterns of interactive online support required to facilitate the user's life/work-style patterns [28-33].

The above priority levels are formally elaborated as follows:

Mandatory – These are those ICT design features which are perceived by the majority of a user group as offering the most needed added-value(s) and that can be accommodated by the target system. This category is expected to include the functionalities that are the common core to all the usage-contexts within the target usage spectrum; including features supporting scalability, modularity and open design so as to enable the incremental evolution of the system to offer further features to satisfy future requirements and customisation as appropriate. They include both selected functional and selected non-functional requirements.

Desirable – These are those features that are desirable, but not highest priority design features as candidates to be accommodated as far as possible, within the resources and technological constraints appertaining to the lifecycle of the project.

Optional – These are those features said by some users to be of the lowest priority and/or are anyway highly contextualised to particular (sub)-sectors of the user group and as such falling into the less common, and/or possibly more controversial and conflictual category.

Once the raw user-stated requirements are aggregated from all elicitation channels and modalities, they have to be transcribed, tabulated and cross-checked to prune duplications and delete clearly out-of-scope requirements. UI-REF promotes a negotiation-based resolution of requirements into the three categories to reflect the priorities of the majority of users [20, 21, 27].

Next additional checks have to be done to flag up for negotiation with the stakeholders the possible deletions, demotions, promotions and new additions of specific requirements to be consensually resolved into the set of mandatory, desirable and optional requirements for the first prototype [20, 21, 27]. The need for the following refinement steps arises as a natural consequence of the fact that the users in stating their requirements can not be expected to be either exhaustive or factor in technology, market and practice constraints (SoA, SoM, SoP) and trends of which they are not necessarily expected to be fully aware. Further, users are expected to articulate their own perceived requirements which may or may not be complete and may be incompatible with other users' requirements or project resources or in conflict with the technological and/or market imperatives and trends [19, 25-27].

Accordingly it will next be necessary to "*factor in*" the influence of the push-pull forces and their dynamics over the near to medium term to ensure that the target system to be delivered will represent the highest rate-of-return on investment for all stakeholders in order to

[1] e.g. by parent sponsor, parent modifier, others etc.

[2] e.g. in writing, video, nested video, emails, on 10th Feb 07 at 11.00 a.m. in UN HQ

have the highest chance of take-up and widest diffusion, usability and technology convergence potential given the current and emergent technological environment that it will have to integrate with i.e. it will be as scalable and sustainable as possible. Such pull and push factors representing constraints and affordances invoked as requirements filters and augmenters can be best understood by performing respectively a SoA, SoP and SoM analysis, where the *State-of-X* is represented by the latest update on the state of current modus operandi, gaps, and, available enabling and emergent innovations from the viewpoint of X. This would serve to indicate the best point of departure for the innovation developments to be achieved within the project, as well as clarify what technologies are needed by the project possibly as COTS that shall not actually be available from external technology in time to be integrated into the project and therefore have to be developed by the project within the allocated resources. Clearly trade-offs have to be considered if there is a project need from external technology that can not itself be fully met in a timely manner in order to enable the project to deliver the full list of user-demanded functionalities within given resources. Accordingly at this stage, by reference to the current SoA, it will be the UI-REF implementers acting as advocates of the system who will need to:

a) ensure that if any of the user-stated requirements imply unrealistic innovation in terms of the delta between current state-of-the art and the innovation required to deliver all the functionalities stated as user requirements, then such a requirement is marked up to be negotiated with the stakeholders as a possible candidate for the lower priority classes of requirements i.e. it should not be classified as a mandatory feature of the target system; at least not for the first prototype stage.

b) flag up for demotion or deletion those requirements that are non-convergent with the current and emergent technology or based on a user-perceived need to integrate the target system at a technically inappropriate level with some obsolete legacy systems. These items would either assume a technology/marketplace paradigm shift that lies beyond the control of the project stakeholders or they would introduce an avoidable convergence problem i.e. sub-optimality of the system in terms of its immediate and future integration potential with the prevailing technology environment.

The ***State-of-the-Market*** (SoM) is represented by the latest update on the relevant products that are available in the marketplace, the relevant gaps in the market that offer synergies with the envisioned functionalities of the target system such that the spotted gap offers an opportunity to add to the exploitation potential of the target system; the SoM is expected to also flags up technology that has already reached market maturity and thus segments that are already approaching the saturation zone thus becoming prohibitive for new entrants. By reference to the current and emergent SoM the UI-REF implementers will be able to suggest:

a) demotion or deletion of any requirements that imply a feature to be introduced that would mean that the resulting system cost or complexity would fall beyond what the market would expect or could possibly tolerate.

b) demotion or deletion of any requirements that imply that a feature should be introduced that would amount to a regressive step in that it has already been tried by others and has either failed outright or proven unsustainable for various reasons and been surpassed by other technology.

c) possible inclusion of any functionality that represents a gap-filler in the relevant products but has not been mentioned in the user-stated requirements list. The possible inclusion of these in the requirements set even as a lower priority feature is ultimately up to the stakeholders but may be suggested as optional if this consideration during the framework architecture specification is deemed to enhance the chances of the target system being capable of integrating such a feature at some point even beyond the lifecycle of the project thus increasing its chances of convergence, take-up, diffusion, scalability, sustainability.

UI-REF, as an evolutionary methodology for requirements engineering and co-design, advocates that the above filters/augmenters are periodically revisited. This is necessary so as to keep up with market and technology evolution through a market-technology-watch task. This task has to report any further filtering/augmenting or other modification suggestions in time to be integrated with the usability evaluation results for the current prototype so as to conclude the requirement engineering update that shall inform the re-engineering and refinement of the next prototype. Thus the application of the above Filters/Augmenters will ensure that additional negotiation is undertaken to resolve those user-stated requirements that are unachievable within the scope of the project and/or will militate against the market trends or mean that obvious current and emergent gaps in the market will not be addressed.

7. The 2020 3d Media project context

The media industry knows that astonishing the public is still a route to large audiences and financial success. It is believed that high quality presentation of stereoscopic or immersive images either in a home or in public entertainment spaces (such as cinemas) can offer previously unimagined levels of experience. Yet this is an almost virgin environment to explore both as far as professional users (producers) and consumers are concerned.

The 2020 3D Media project research and technology innovating programme aims to demonstrate novel forms of compelling entertainment experiences based on new technologies for the capture, production, networked distribution and display of three-dimensional sound and images.

Target users are expected to be mainly media industry professionals across the current film, TV and 'new media' sectors producing programme material not disregarding the general public that in the end will consume such products. The potential advantages include:

- Heightened reality and a renewed sense of presence, putting the spectator at the heart of a more exciting experience.
- The ability for the spectator to navigate a virtualised world that has a complete sense of reality.
- The ability to change things in this world once it has been created.
- The ability to repurpose and deploy multi-dimensional content in different contexts

Yet a major problem is that there is not yet a well defined and formalised to serve multi-agent communication and coordination as well as the orchestration of agent workflows for shooting and making 3D products suitable for the mass market, or at least nothing comparable to what is currently available for the television and film industry.

The techniques available for collaborative production coordination in the 3D environments are still much in the experimental phase and have yet to leave the R&D laboratories. Some pioneering authors (e.g. Warner Bros. *House of Wax* in 1953) have made some significant contributions; however the research to establish a robust, efficient and scalable multi-agent coordination expression language to serve the domain of collaborative 3D media production is still in progress.

8. UI-REF for 2020 3D Media

To overcome the challenges in collaborative workflow support in the 3D production domain, the underlying semantics of Storyboarding, Communications and Control (SCC) have to be modelled so as to provided legacy-workflow-agnostic support for integrated multi-agent coordination in 3D production

This will essentially involve a framework for abstract workflow representation that could instantiate organised networks of basic-component work-flows modelling the various and heterogeneous process combinations that produce the expected 3D media product. The UI-REF framework has been deployed to capture the user requirements for such a supportive platform..

In the context of 2020 3D Media we have already conducted the pre-requisite Domain Knowledge Analysis based on reference material describing the domain(s) – namely cinema, television, (post)-production, 3D, stereoscopy, etc. – addressed directly or indirectly by analysing integrated multi-agent workflows or by observations made through inter views with practitioners re domain ontology, practice logics, work-styles and project activities etc. This has yielded the initial basis for the relevant ontology i.e. glossaries of the domain terminology descriptions to serve as the basis of shared understanding to support the subsequent requirements elicitation. This is an integrated glossary (REF website) that facilitates mutual understanding of the knowledge basis for the domain workflow models and the user requirements re the target virtualisation environment to support users in multi-level SCC during collaborative working on 3D productions.

We have conducted research into state-of-the-art for modelling and supporting the workflow logic in cinema and television production [83-95]. This has led to a basic understanding of the problem, its boundaries, and situated usage-contexts involving various actors, processes, objects, states and triggers. Once all these had been elicited, we conducted a series of structured interviews with selected users to collect their visions of the target solution, their requirements and needs and further enrich the picture that was emerging from the context, market and technology analysis.

Results of such meetings (in terms of minutes) have been collected in documents shared with partners to elicit discussion and feedback.

9. Ongoing and future work

At this stage, the information elicited through the process so far has been analysed and further interviews are planned are planned to explicated the contexts of usage to be distinguished within the end-to-end cooperation patterns of interacting workflows due the distinct requirements subsets that may become dominant in or be exclusive to each phase of the creative process from conceptualisation to fruition.

The final result will be a document that describes the various usage-contexts, respective (sub)ontologies, the underpinning domain ontology of actors, objects, processes, triggers, etc; it will also provide the generalisation ontology f domain usage contexts and requirements. Subsequently the steps elaborated in UI-REF reference document [95] and summarised in sections 1 to 9 of this document will be deployed to arrive at an initial prioritisation of the usage-context-specific requirements. An example of this is illustrated in Fig 1, as a small segment of a requirements resolution table as derived in the context of a Computer Network Security project [17]. In this example, the first three columns identify and set out the questionnaire-root (the question that was responsible for eliciting the requirement element), the following columns set out the initial and finally resolved form of the requirements hierarchy for each prioritised and respectively colour-highlighted requirement element.

In UI-REF such an initial requirements prioritisation is to be negotiated with the user groups so as to conclude the final resolution of the domain requirements priorities.

PART	SECTION	No	QUESTION	Conclusion from Responses	Finalised Requirements Hierarchy
A	Organisa-tional Profiles	1	This section attempted to clarify the organisational setting in which the network security executive (the interviewee) had to operate in terms of configuration, policy sectorial-specifics etc	The Network Security Management Context of the organisation being interviewed was clarified.	The Network Security Management Context of the organisation being interviewed was clarified.
B.	General Security Concerns	1.	What is your view on past innovations & significant future/promising technologies (e.g.: bio-inspired ones)?	M: Filters updates, support in effective security policy and threats management, real-time monitoring & assessment of risk O: Very good filtering of spam & other threats.	M: Implement & update of high-quality filters. M: Support- effective security policy D: Provide a real-time view of risk O: Managing security problem.
		2.	Attack Detection – what dangers do you most want to be defended against?	M: Attack detection, alerting of malware D: Protect against DoS, virus, spam, phishing. O: Enhanced identification	M: Detection of Malware, DoS, Virus , SPAM O: Better visibility of threats O: enhanced identification
		...	...	...	...

Fig. 1 – Excerpt of the Requirement resolution table of FastMach Project [17]

10. Conclusions

The proposed methodology framework for user-intimate requirements hierarchy resolution (UI-REF) has been based on well established and grounded references at international level and complemented with the experience and research of authors in the usage of MIL-STD-498 and MIL-STD1472F, MMREA [13], PopEval-MB / WebEval-AB [15] and simple office automation tools and templates for an easy implementation in most application environments dealing with media and systems.

The present paper encapsulates the salient precepts of the UI-REF methodology framework [95] and sets this in the context of a current UI-REF deployment for requirements engineering for ICT support for multi-agent collaborative storyboarding, communications and control of workflow in 3D Media (Post)-Production & Distribution.

11. References

[1] *Guide to the Software Engineering Body of Knowledge SWEBOK v0.9* IEEE Computer Society & ACM 2001

[2] *MIL-STD-498 Software Development and Documentation Std*, US Department Of Defence, 1995

[3] *MIL-STD-1472F Human Engineering*, US Department of Defence, 1999

[4] Maguire M C, *RESPECT User-Centred Requirements Handbook, Version 3.3*, RESPECT Project 1998

[5] *A guide to the Project Management Body of Knowledge 3rd Edition (PMBOK® Guide) – an American National Standard ANSI/PMI 99-001-2004*, Project Management Institute Inc., 2004

[6] *PRINCE 2 Pocketbook* by OGC - Office of Government Commerce, 2002

[7] Pyzdek T., *The Six Sigma Handbook*, Mc Graw Hill, 2003

[8] Bentley C., *Prince 2: A Practical Handbook*, Butterworth-Heinemann, 2002

[9] Tennant G., *Six SIGMA: SPC and TQM in Manufacturing and Services*, Gower Publishing, Ltd, 2001

[10] van Bon J., Pieper M., van der Veen A., *Foundations of IT Service Management Based on ITIL: Based on ITIL*, itSMF - IT Service Management Forum, 2005

[11] *Introduction to ITIL*, The Stationery Office, 2005

[12] *Merriam-Webster's Collegiate Dictionary & Thesaurus*, Merriam-Webster Incorporated, 2003

[13] Hounat H., Badii A., "Multi-media Requirements Elicitation Assistant (MMREA)"; proceedings of the 1st International Conference on Knowledge Transfer (KT'96), London University, July 1996.

[14] Eva M., Badii A., "What will it do to us? Predicting the cultural impact of new ICT applications deployment"; proceedings of the 3rd Projectics Conference, Bayonne-Cedex, December 1997.

[15] Badii A., "Online point-of-click web usability mining with PopEval-MB, WebEval-AB and the C-Assure methodology"; proceedings of Americas Conference on Information Systems (AMCIS 2000), University of California, Long Beach, August 2000.

[16] Turk A, Badii A, "Personalized, Mediated Human-Computer Interaction", Proceedings of the Western Australian Information Systems Conference July 2001.

[17] "FastMatch Network Security Framework architecture Requirements", FastMatch Project

[18] Bryne E., "IEEE Standard 830: Recommended Practice for Software Requirements Specification," IEEE International Conference on Requirements Engineering, IEEE Computer Society Press, April 1994, p. 58.

[19] Davis A.M., "Software Requirements: Objects, Functions and States". Prentice-Hall, 1993.

[20] Goguen J., and Linde C., "Techniques for Requirements Elicitation", International Symposium on Requirements Engineering, San Diego, California: IEEE Computer Society Press, January 1993, pp. 152-164.

[21] Kotonya G., and Sommerville I., *Requirements Engineering: Processes and Techniques*. John Wiley and Sons, 2000.

[22] Loucopulos P., and Karakostas V., *Systems Requirements Engineering*. McGraw- Hill, pp. 1995.

[23] Pfleeger S.L., *Software Engineering- Theory and Practice*. Prentice-Hall, Chap. 4, 1998.

[24] Rosenberg L., Hammer T.F. and Huffman L.L., "Requirements, testing and metrics", 15th Annual Pacific Northwest Software Quality Conference, Utah, October 1998.

[25] Sommerville I., *Software Engineering (6th edition)*, Addison-Wesley, pp. 63-97, 97-147, 2001.

[26] Sommerville I., and Sawyer P., *Requirements engineering: A Good Practice Guide*. John Wiley and Sons, Chap. 1-2, 1997.

[27] Thayer R.H., and Dorfman M., *Software Requirements Engineering (2nd Ed)*. IEEE Computer Society Press, pp. 176-205, 389-404, 1997.

[28] Badii A., Sailor M., Marangoni C., Neglia A., and Pearce L., "Workflow Data Exchange Semantics, Object Discovery & Integration Architecture for Cross Media Production & Distribution", 1st AXMEDIS International Conference, Florence, Italy, 30 Nov – 2 Dec 2005.

[29] Badii A., Sailor M., Marangoni C., Neglia A., and Pearce L., "Object Life Cycles & Interaction Environments for Cross Media Production & Distribution", 1st AXMEDIS International Conference, Florence, Italy, 30 Nov – 2 Dec 2005.

[30] Ng K., Badii A., Sailor M., Ong B., Neagle R., Quested G., "AXMEDIS Programme and Publication Tools Integration with Workflow-enabled Communication and Process Control". In Proceedings of the 2nd AXMEDIS International Conference, University of Leeds, UK, 13-15 Dec. 2006. Pages 110-120

[31] Badii A., Sailor M., Nair R., "Profiling Management for Personalised Multimedia Delivery On-Demand within the AXMEDIS Framework", In Proceedings of the 2nd AXMEDIS International Conference, 2006. AXMEDIS '06. Page 35-44, IEEE Press, IEEE, Leeds, UK, December 2006

[32] Badii A., Khan A., Adetoye A., Fuschi D., "High Resolution Digital Media Personalisation for All Delivery Context Device Agents". In: The 14th International Conference on Distributed Multimedia Systems (DMS '08), 4-6 September 2008, Boston, USA

[33] Badii A., Khan A., Adetoye A., Fuschi D., "Personalised Digital Media Adaptation and Delivery to Mobile Devices". In: Third International Workshop on Mobile and Networking Technologies for Social Applications (MONET′08), Nov 9 - 14, 2008, Monterrey, Mexico

[34] Teichrow D. and Hershey E.A., "PSL/PSA: A Computer Aided Technique for Structured Documentation and Analysis of Information Processing Systems," *IEEE Trans. Software Eng*., vol. SE-3, no. 1, 1977, pp. 41–49.

[35] DeMaro T. and Plauger P.J., *Structured Analysis and System Specification*, Prentice Hall, 1979.

[36] Jacobson, *Object-Oriented Software Engineering: A Use-Case Driven Approach*, Addison-Wesley, 1992.

[37] Booch G., Rumbaugh J., and Jacobson I., *The Unified Modeling Language User Guide*, Addison-Wesley, 1998.

[38] IEEE Std IEEE-Std-830-1998, *IEEE Recommended Practice for Software Requirements Specification*, IEEE CS Press, 1998.

[39] Sommerville and Sawyer P., "Viewpoints: Principles, Problems and a Practical Approach to Requirements Engineering," *Annals of Software Eng*., vol. 3, 1997, pp. 101–130.

[40] Hall, "Using Formal Methods to Develop an ATC Information System," *IEEE Software*, vol. 13, no. 2, 1996, pp. 66–76.

[41] Van Lamsweerde, "Goal-Oriented Requirements Engineering: A Guided Tour," *Proc. 5th Int'l IEEE Requirements Eng. Conf.*, IEEE CS Press, 2001, p. 249.

[42] Sommerville and Sawyer P., *Requirements Engineering: A Good Practice Guide*, John Wiley & Sons, 2000.

[43] Davis, *Software Requirements: Analysis and Specification*, Prentice Hall, 1990.

[44] Beck K., *Extreme Programming Explained: Embrace Change*, Addison-Wesley, 2000.

[45] Boehm et al., "Requirements Engineering, Expectations Management and the Two Cultures," *Proc. 7th Int'l Symp. Requirements Eng.*, IEEE CS Press, 1999, pp. 14–22.

[46] Boehm et al., "Using the WinWin Spiral Model: A Case Study," *Computer*, vol. 31, no. 7, 1998, pp. 33–44.

[47] Maiden N.A. and Ncube C., "Acquiring COTS Software Selection Requirements," *IEEE Software*, vol. 15, no. 2, 1998, pp. 46–56.

[48] Boehm B. and Abts C., "COTS Integration: Plug and Pray?" *Computer*, vol. 32, no. 1, 1999, pp. 135–138.

[49] Lausen S., "COTS Tenders and Integration Requirements." *Proc. 12th IEEE Int'l Requirements Eng. Conf.*, IEEE CS Press, 2004, pp. 166–175.

[50] Hofmann H.F., *Requirements Engineering: A Situated Discovery Process*, Gabler, Wiesbaden, Germany, 2000.

[51] *IEEE Guide to Software Requirements Specification*, IEEE Std. 830-1998, IEEE Press, Piscataway, N.J., 1998.

[52] Jones, *Applied Software Measurement: Assuring*

Productivity and Quality, McGraw-Hill, New York, 1996.
[53] Macaulay L.A., *Requirements Engineering*, Springer, London, 1996.
[54] Chatzoglou P.D., "Factors Affecting Completion of the Requirements Capture Stage of Projects with Different Characteristics," *Information and Software Technology*, vol. 39, no. 9, Sept. 1997, pp. 627–640.
[55] Curtis, Krasner H., and Iscoe N., "A Field Study of the Software Design Process for Large Systems," *Comm. ACM*, vol. 31, no. 11, Nov. 1988, pp. 1268–1287.
[56] Emam K E and Madhavji N H, "A Field Study of Requirements Engineering Practices in Information Systems Development," *Second Int'l Symp. Requirements Eng.*, IEEE CS Press, Los Alamitos, Calif., 1995, pp. 68–80.
[57] Doherty N.F. and King M., "The Consideration of Organizational Issues during the System Development Process: An Empirical Analysis," *Behavior & Information Technology*, vol. 17, no. 1, Jan. 1998, pp. 41–51.
[58] Guinan P.J., Cooprider J.G., and Faraj S., "Enabling Software Development Team Performance During Requirements Definition: A Behavioral Versus Technical Approach," *Information Systems Research*, vol. 9, no. 2, 1998, pp. 101–125.
[59] Lubars M., Potts C., and Richter C., "A Review of the State of the Practice in Requirements Modeling," *First Int'l Symp. Requirements Eng.*, IEEE CS Press, Los Alamitos, Calif., 1993, pp. 2–14.
[60] Nidumolu S.R., "A Comparison of the Structural Contingency and Risk-Based Perspectives on Coordination in Software-Development Projects," *J. Management Information Systems*, vol. 13, no. 2, 1995, pp. 77–113.
[61] *IEEE Guide to Software Requirements Specification*, IEEE Std. 830-1998, IEEE Press, Piscataway, N.J., 1998.
[62] H.F. Hofmann, *Requirements Engineering: A Situated Discovery Process*, Gabler, Wiesbaden, Germany, 2000.
[63] Boehm B.W., *Software Engineering Economics*, Prentice Hall, Englewood Cliffs, N.J., 1981.
[64] Jacobson, *Object-Oriented Software Engineering*, Addison Wesley Longman, Reading, Mass., 1992.
[65] Rumbaugh J., "Getting Started: Using Use Cases to Capture Requirements," *J. Object-Oriented Programming*, Vol. 7, No. 5, Sept. 1994, pp. 8–12.
[66] Rowlett T., "Building an Object Process around Use Cases," *J. Object-Oriented Programming*, Vol. 11, No. 1, Mar./Apr. 1998, pp. 53–58.
[67] Dano, H. Briand, and F. Barbier, "Progressing towards Object-Oriented Requirements Specifications by Using the Use Case Concept," *Proc. Int'l Conf. Requirements Eng.*, IEEE Computer Soc. Press, Los Alamitos, Calif., 1996, pp. 450–456.
[68] Rubin K.S. and Goldberg A., "Object Behavior Analysis," *Comm. AC*M, Vol. 35, No. 9, Sept. 1992, pp. 48–62.
[69] Pohl K., "The Three Dimensions of Requirements Engineering: A Framework and Its Applications," *Information Systems*, Vol. 19, No. 3, 1994, pp. 243–258.
[70] Muller P.A., *Instant UML*, Wrox Press Ltd., Olton, Birmingham, UK, 1997.
[71] Regnell B., Andersson M., and Bergstrand J., "A Hierarchical Use Case Model with Graphical Representation," *Proc. IEEE Symp. and Workshop on Engineering of Computer-Based Systems*, IEEE Computer Soc. Press, Los Alamitos, Calif., 1996, pp. 270–277.
[72] Borgida, Greenspan S., and Mylopoulos J., "Knowledge Representation as the Basis for Requirements Specification," *Computer*, Apr. 1985, pp. 82–91.
[73] Finkelstein and Waters R.C., "Summary of the Requirements Elicitation, Analysis and Formalization Track," *ACM Software Eng. Notes*, Vol. 14, No. 5, 1989, p. 40.
[74] Georgeff M P and Lansky A L, "Procedural knowledge." *Proc. IEEE*, Vol. 74, No. 10, Oct. 1986, pp. 1383–1398.
[75] Potts, Takahashi K., and Anton A.I., "Inquiry-Based Requirements Analysis," *IEEE Software*, Mar. 1994, pp. 21–32.
[76] Dardenne, van Lamsweerde A., and Fickas S, "Goal-Directed Requirements Acquisition," *Science of Computer Programming*, Vol. 20, 1993, pp. 3–50.
[77] Mylopoulos J., Chung L., and Nixon B., "Representing and Using Nonfunctional Requirements: A Process-Oriented Approach," *IEEE Trans. Software Eng.*, Vol. 18, No. 6, 1992, pp. 483–497.
[78] Lee J. and Kuo J.Y., "New Approach to Requirements Trade-OffAnalysis for Complex Systems," *IEEE Trans. Knowledge and Data Eng.*, Vol. 10, No. 4, July/Aug. 1998, pp. 551–562.
[79] Cockburn, "Goals and Use Cases," *J. Object-Oriented Programming*, Vol. 10, No. 7, Sept. 1997, pp. 35–40.
[80] Boehm and In H., "Identifying Quality-Requirement Conflicts," *IEEE Software*, Mar. 1996, pp. 25–35.
[81] van Lamsweerde, Darimont R., and Massonet P., *Goal-Directed Elaboration of Requirements for a Meeting Scheduler Problems and Lessons Learnt*, Tech. Report RR-94-10, Universite Catholique de Louvain, Louvain-la-Neuve, Belgium, 1994.
[82] Karlsson J. and Ryan K., "A Cost-Value Approach for Prioritizing Requirements," *IEEE Software*, Sept./Oct. 1997, pp. 67–74.
[83] Ouyang C., La Rosa M., ter Hofstede A.H.M., Dumas M., Shortland K.; "Towards Web-ScaleWorkflows for Film Production"; 2008; QUT Digital Repository (ID 13298.pdf)
[84] Ouyang C., ter Hofstede A.H.M., La Rosa M., Rosemann M., Shortland K., Katherine and Courtavid D.; "Camera, Set, Action: Automating Film Production via Business Process Management"; 2008; QUT Digital Repository (ID 13361.pdf)
[85] van der Aalst W.M.P., and ter Hofstede A.H.M.; "YAWL: Yet Another Workflow Language" *Infor-*

mation Systems, 30(4): 245–275; 2005.

[86] van der Aalst, W.M.P., ter Hofstede, A.H.M., and Weske, M. "Business Process Management: A Survey". In *Proceedings of the Business Process Management 2003*, volume 2678 of Lecture Notes in Computer Science, Springer-Verlag, pp.1-12. 2003.

[87] Clevé, B. *Film Production Management*. Burlington, Oxford. 2006.

[88] Hevner, A.R., March, S.T, Park,J., and Ram, S. "Design Science in Information Systems Research". *MIS Quarterly*, 28(1): 75-105. 2004.

[89] van der Aalst W.M.P,. ter Hofstede A.H.M, Kiepuszewski B., and Barros A.P. "Workflow Patterns". *Distributed and Parallel Databases*, 14(1):5–51, 2003.

[90] Blake M.B. and Huhns M. *Web-scale workflows: Integrating distributed service*". IEEE Internet Computing, 12, January 2008.

[91] Holmes T., Vasko M., and Dustdar S.. "VieBOP: Extending BPEL Engines with BPEL4People", *16th IEEE Euromicro International Conference on Parallel, Distributed and Network-based Processing*, Toulouse, France, February 2008.

[92] Jordan D.and Evdemon J. *Web Services Business Process Execution Language Version 2.0.* Committee Specification. OASIS WS-BPEL TC, January 2007. Available via http://www.oasis-open.org/committees/download.php/22475/wsbpel-v2.0-CS01.pdf.

[93] La Rosa M., Lux J., Seidel S., Dumas M., and ter Hofstede A. H. M. "Questionnaire-driven Configuration of Reference Process Models" *Proceedings of the 19th International Conference on Advanced Information Systems Engineering (CAiSE'07)*, Trondheim, Norway, 11-15 June 2007.

[94] Rosemann M.and van der Aalst W.M.P. "A Configurable Reference Modelling Language". *Information Systems*, 32(1):1–23, 2007.

[95] Badii, A: *UI-REF Reference Framework for User-Intimate Systems Requirements Engineering based on C-Assure Methodolgy, integrating Dynamic usability requirements modelling-evaluation, and, virtualisation-enabled use-centric co-deign CompanionAble Project Deliverable May 2008*., IMSS Technical Report, 1-1-05-08, School of systems Engineering University of reading.

[96] Badii A, Sharif A, "Information management and knowledge integration architecture for enterprise innovation", *International Journal of Logistics Information Management*, Vol 16, No. 2, 2003, pp 145-155 (Best Paper of the Year Award).

[97] Badii A, C-Assure-Taguchi framework for cost-effective holistic IS evaluations, International Journal of Enterprise Information Management, special issue on information systems evaluation, Vol 17, No 4, 2004, pp 304-314

[98] Badii A, Context, context, wherefore art thou? shared sense-making and meta-scaffolding for collaborative creative communities; Proceedings of the 1st UKAIS-AIS Workshop on Interpretive Approaches in Information Systems Research, Brunel University, 25-27 July 2002.

[99] Parsons-Hann H, Liu K, Badii A, Using The Semiotic Framework To Derive A Generic Virtual Organisational Architecture, Proceedings Of The EMCIS '04, 25-27 July 2004.

[100] Badii A, Information System Integration, a tutorial paper, Proceedings of the 6th United Kingdom Academy of Information Systems Conference (UKAIS 2001), Portsmouth University, April 2001.

[101] Badii A, A framework for Cultural Accommodation with Sensitised Systems for Usability Relationships Evaluation (C-Assure) in abstract Proceedings of the United Kingdom Academy of Information Systems Research Seminars Series 1999, UKAIS Newsletter, June 2000

[102] Zhang W, Badii A, Integrated object oriented architecture for workflow-embedded document management, Proceedings of the 2nd International Conference on Knowledge Transfer (KT'97), University College London, July 1997.

[103] Badii A, Integrated Data Discovery, Exchange Engineering and Mining Management Architecture (ID-DEEMA) using communities of foraging agents; Proceedings of 1st. International Conference on Knowledge Transfer (KT'96), London University, July 1996.

[104] Badii A, Rolfe R, Boundary sensitised (electronic) relationship management; Proceedings of the 1st UK Academy of Information Systems Conference (UKAIS'96), Cranfield University, April 1996.

AUTHOR INDEX

AXMEDIS 2008: proceedings of the 4th International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution. Workshops, Panels and Industrial Applications. Florence, Italy 17-19 November 2008
edited by Paolo Nesi, Kia Ng, Jaime Delgado
ISBN: 978-88-8453-811-6 (online). ISBN 978-88-8453-810-9 (print).

Finito di stampare presso
la tipografia editrice Polistampa