

Da Big Data Analytic ai Data Models e Digital Twin

Paolo Nesi

Nei primi anni del 2000 la generazione di dati è cresciuta in modo esponenziale e con questa la necessità di gestirli e l'opportunità di produrre conoscenza da questi. Uno dei fattori che hanno provocato questa accelerazione è stata la spinta europea agli *Open Data* per una amministrazione trasparente e per arrivare a processi decisionali guidati dai dati (data-driven) a vari livelli sia del decisore che del consumatore (utente finale del servizio). Tale impulso ha spinto a rendere pubblici dati statistici e puntuali dei singoli comuni, province, regioni fino alle grandi istituzioni e alle grandi industrie. Fra questi dati sono comparsi inizialmente molti dati statistici, in seguito quelli satellitari di varia natura, una miriade di dati relativi a misurazioni puntuali dal mondo dell'Internet delle Cose (Internet of Things, IoT), come dati da sensori ambientali, di traffico, inquinamento, mezzi di trasporto, energia, ecc. Inoltre, i flussi dati prodotti dal comportamento delle persone, dai loro spostamenti, dagli acquisti, il comportamento delle persone tramite i tracciati della telefonia mobile, da telecamere, dai contapersone, dalle scatole nere delle assicurazioni sui mezzi privati, e sui mezzi di trasporto pubblici, ecc. (Internet of Everything, IoE). Ma anche i dati dei social media, i post sui blog, i tweet, ecc.

In tali anni, il DISIT lab dell'Università di Firenze ha iniziato ad estendere la sua attività nel campo dei Big Data e del Big Data Analytic abbandonando progressivamente la tematica dei media che aveva prodotto rilevanti progetti di ricerca AXMEDIS, MO-ODS, Wedelmusic, ecc., per andare a trattare dati molto più eterogenei e complessi. In questo nuovo contesto, partendo dalla definizione dei modelli e delle tecniche di intelligenza artificiale per l'analisi del testo ed il ragionamento automatico logico/semantico, ed in seguito con le previsioni su serie storiche, l'identificazione di anomalie (diagnosi precoce, *early warning*), per arrivare allo sfruttamento dei dati per i sistemi di supporto alle decisioni con approcci neuro-simbolici. Questi erano da tempo temi di ricerca di base all'interno del DISIT lab su progetti tipicamente in ambito industriale, beni culturali e media dove si generavano e si dovevano manipolare grandi quantità di dati. Il mondo degli open data, dell'IoT/IoE fino al Web of Things, WoT, hanno aperto l'uso di tecniche

Paolo Nesi, University of Florence, Italy, paolo.nesi@unifi.it, 0000-0003-1044-3107

Referee List (DOI 10.36253/fup_referee_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Paolo Nesi, *Da Big Data Analytic ai Data Models e Digital Twin*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0975-5.32, in Stefano Selleri, Alberto Tesi, Enrico Vicario (edited by), *Ingegneria Industriale & Ingegneria dell'Informazione per il territorio fiorentino – 2. Ingegneria dell'Informazione*, pp. 127-131, 2026, published by Firenze University Press, ISBN 979-12-215-0975-5, DOI 10.36253/979-12-215-0975-5

di Big Data Analytic/Artificial Intelligence e dei Sistemi Distribuiti ai sistemi con alta densità di dati e processi, connessi e controllati da interventi decisionali umani come le Smart City, i processi decisionali industriali i sistemi, i sistemi CyberPhysical, e i sistemi Cloud. Nasce in quegli anni la soluzione ICARO cloud (2015) come risultato dell'omonimo progetto di ricerca, in collaborazione con Var Group per la gestione intelligente di grandi data center in cloud, e cloud service provider, per l'ottimizzazione delle risorse ed allocazione automatica dei processi e delle applicazioni gestite in cloud di grandi dimensioni, tramite tecniche inferenziali e ragionamento automatico sull'ontologia smart cloud definita dal DISIT Lab, soluzioni di semantic computing.



Figura 72 – La dashboard primaria della control room della Città di Firenze con il Sindaco Dario Nardella (2019).

Una pietra miliare del DISIT lab è stata la realizzazione del modello Km4City (Knowledge Model for the City) presentato nel 2013 nel Salone dei Cinquecento a Palazzo Vecchio, Firenze. Era la base di quello che è diventato un sistema esperto per la città che sfrutta la modellazione dei dati a livello cittadino e regionale, e il ragionamento inferenziale semantico, oggi si direbbe simbolico. Il primo impiego si è visto nell'aggregatore dati per la Città di Firenze, ed in seguito come parte centrale del progetto Sii-Mobility di ricerca Smart City nazionale per la mobilità ed i trasporti sostenibili, il progetto nazionale del Ministero dell'Istruzione, dell'Università e della Ricerca, MIUR, su mobilità e trasporti ideato nel 2012 e operativo fino al 2022, coordinato da DISIT lab, che ha prodotto la piattaforma Sii-Mobility open source per il riuso della pubblica amministrazione, ma anche modelli per la ricostruzione del traffico da equazioni differenziali, modelli predittivi su parcheggi e bike sharing (con soluzioni di deep learning, e di Explainable AI, XAI), analisi di dati dai social media (con tecniche di NLP, natural language processing, e sentiment analysis, con la soluzione Twitter Vigilance per l'analisi social media), ecc., e sperimentazioni su tutta la Toscana coinvolgendo operatori del trasporto pubblico locale e nazionale come: ATAF, BUSITALIA, CTTnord, Tiemme, ECM, e molte altre aziende del territorio fiorentino e toscano, e decine di pubblicazioni internazionali e soluzioni e prodotti derivati. La ricerca in Sii-Mobility ha sviluppato il Sistema Integrato Interoperabile, SII, con alla base Km4City, per le smart city e la mobilità utilizzato in seguito alla base di progetti europei di ricerca di grandi dimensioni come REPLICATE H2020 e RESOLUTE H2020. REPLICATE è stato un progetto faro smart city della Commissione Europea, nel quale è stata realizzata la prima versione della Smart City Control Room di Firenze, tutt'ora attiva ed in uso del Sindaco Dario Nardella (2023). Il progetto RESOLUTE di ricerca ha sfruttato la modellazione Km4City ed il motore Big Data Analytic per la definizione delle linee guida Europee per la resilienza delle infrastrutture critiche, con attenzione ai sistemi

di trasporto in Firenze ed Atene, tenendo conto dei dati del traffico, dell'inquinamento e dei flussi delle persone, e per la sua realizzazione operativa fino al computo automatico del sistema di supporto alle decisioni. Con questi progetti, il modello Km4City ed il SII si sono evoluti fino a fornire soluzioni a supporto delle decisioni, anche a fronte di eventi inattesi, gli *unexpected unknowns*. Per esempio, la bomba d'acqua in Lungarno Colombo.

Vengono istituiti in quegli anni corsi Universitari per la Magistrale come Big Data Architecture, Knowledge Engineering e anche il master MABIDA in Big Data Analytics and Technologies for Management dell'Università di Firenze.

La ricerca del DISIT lab in ambito Big Data Analytic, Smart City ed IoT si è spinta ancora maggiormente a livello internazionale con il progetto Snap4City e l'omonima piattaforma¹, che ha visto il DISIT lab cimentarsi in una competizione internazionale (Select4Cities) finanziata dalla Commissione Europea e diretta da Copenaghen, Antwerp ed Helsinki per la selezione della migliore soluzione in ambito smart city ed IoT/IoE. Il consenso su Snap4City è cresciuto notevolmente anche grazie al modello a MicroServizi, ed al suo sviluppo in termini di oltre 180 nodi e 4 libraries per la piattaforma Node-RED della JS Foundation. Risultato: Snap4City ha ricevuto l'Award del vincitore di tale competizione nel 2019 a Barcellona, allo Smart City Expo World Congress, presentando una soluzione capace di integrare modelli smart di visualizzazione Big Data Analytic, un motore di intelligenza artificiale (machine learning e deep learning e XAI), un modello Km4City avanzato, una soluzione integrata basata su MicroServizi per la programmazione visuale di processi smart, l'assistente virtuale conversazionale PAVAL in grado di rispondere a domande su Firenze e dare indicazioni turistiche con tecniche di NLP, completamente open source, conforme all'allora nuova normativa sulla privacy, il GDPR. La stessa soluzione Snap4City è stata in seguito certificata dalla fondazione FIWARE (CEF della Commissione Europea), è presente su EOSC (European Open Science Cloud), è presente in E015 regione Lombardia, ed in trial presso il JRC ISPRA della Commissione Europea.

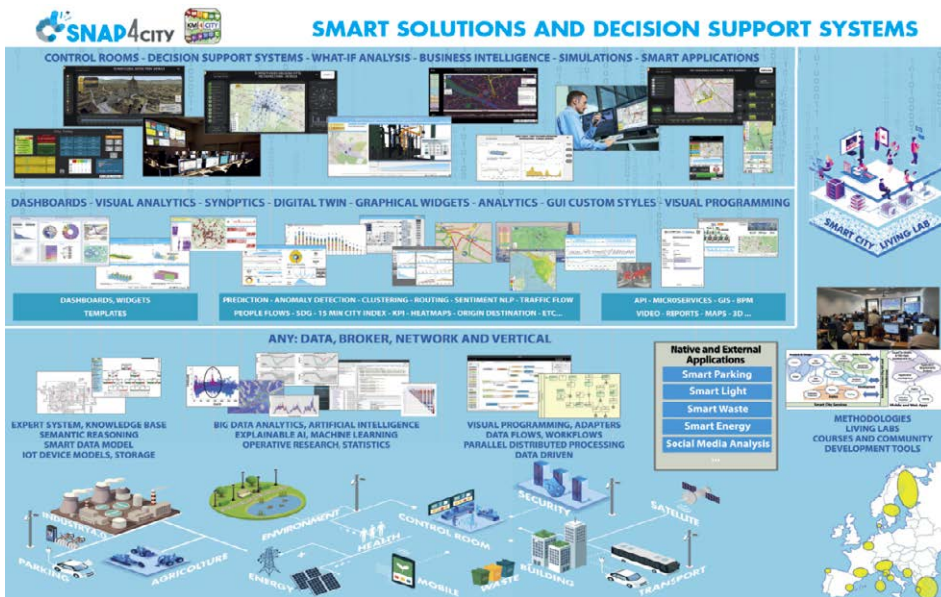


Figura 73 – La sintesi della piattaforma Snap4City.

¹ <https://www.Snap4City.org>

Si leggeva in quel periodo: «1st place has been awarded to Florence (Snap4City) [...] the competition was really hard [...] about Snap4City strong points as always, excellent analytics, excellent data aggregation».

Negli stessi anni nasceva la Città Metropolitana di Firenze, una delle dieci città metropolitane italiane. Il DISIT lab ha coordinato la parte ICT del piano strategico della Città Metropolitana di Firenze. In seguito, questa capacità e competenza del DISIT lab per la definizione di piani strategici per le Smart City si è confermata con lo sviluppo di quello del ministero dell'innovazione di Cypro, dove il DISIT lab ha coordinato le competenze di esperti internazionali e delle Università di Pisa e Cagliari.

Sulla linea di ricerca su Big Data Analytic per le Smart City/IoT con Snap4City si sono innestati a cascata svariati altri progetti. TRAFAIR CEF, progetto di ricerca della Commissione Europea (2018) dove il DISIT lab ha prodotto modelli di ricostruzione del traffico e per le predizioni del NOX (NO, NO2) tramite modelli di Artificial Intelligence e della fluidodinamica nei modelli 3D della città sulla base del traffico in Firenze, Pisa e Livorno, ma anche in altre città europee. TRAFAIR ha vinto il ECOHiTech Award 2022; MOBIMART Interreg per la smart city di Pisa che ha prodotto Snap4Pisa; Life Weee per la raccolta intelligente di RAEE (rifiuti ad elevato valore: Rifiuti di Apparecchiature Elettriche ed Elettroniche) in Toscana e in Catalogna; 5G per la sperimentazione delle tecnologie 5G dove DISIT lab ha contribuito con la piattaforma Snap4City. Nasce il progetto MOSAIC di ricerca con Alstom e altri partner, dove DISIT Lab ha prodotto DORAM per i modelli di analisi a supporto delle decisioni in ambito del trasporto pubblico e per valutare la soddisfazione della domanda di mobilità sulla base dell'offerta di trasporto pubblico (bus, ferrovie leggere, tramvia, ecc.), e la What-If analysis sui servizi di trasporto pubblico (rispondendo a domande del tipo: «Cosa accade al trasporto pubblico se si operano delle modifiche al servizio anche fuori dagli schemi e dalle esperienze pregresse?», quali saranno le fermate critiche in termini di persone in attesa alla fermata?). Il progetto GHOST MIUR SIR a Cagliari per il governo delle smart city in tema di smart governance.

Nel 2019, prima della pandemia COVID-19, DISIT lab ha messo a punto un modello con Snap4City per la valutazione e la pianificazione delle città sulla base di dati accessibili in modo da rispettare il concetto della città a 15 minuti, per il quale ogni area della città dovrebbe fornire servizi accessibili nella distanza di 15 minuti a piedi da dove vivono i cittadini (servizi di mobilità, governo, lavoro, salute, educazione, sport, cultura, culto, ecc.). Il modello, 15 Min City Index del DISIT lab ha vinto la challenge internazionale di ENEL-X dimostrandone la validità sulla Città Metropolitana di Firenze, e fornendo il supporto per poterlo applicare sulla Città Metropolitana di Bologna. In seguito, il modello a 15 Minuti del DISIT è stato adottato da ENEL-X per realizzare un servizio a livello nazionale, operativo dal 2021 su tutte le microaree a 15 minuti degli 8000 comuni italiani. Anche questo ha vinto l'ECOHiTech Award 2022. Nasce in questi anni lo Snap4, *spin-off* dell'Università di Firenze, in seguito trasformato in *spin-off* accademico e start up innovativa.

Nel 2020 è iniziato il progetto di ricerca Herit-Data Interreg della Commissione Europea per comprendere e definire soluzioni basate su dati nel contesto dei flussi turistici in Firenze, Valencia, Pont du Gard, Patrasso, Mostar e Dubrovnik, che sono aree turistiche di grande attrazione in Europa. Il DISIT lab ha utilizzato la piattaforma Snap4City per lo sviluppo di modelli Big Data Analytics, di analisi e predittivi per i flussi di persone tramite dati eterogenei da: sensori, telecamere termiche, arrivi/partenze, social media da Twitter; usando anche tecniche NLP e Sentiment Analysis su Twitter Vigilance. Sono stati prodotti modelli predittivi della reputazione dei servizi e delle presenze nelle aree critiche delle città. Nello stesso contesto, è stato sviluppato

il primo Digital Twin globale della città di Firenze, che fornisce una rappresentazione navigabile del modello 3D di Firenze ed integrata ai dati real time della Control Room di Firenze, ed integrata con alcuni modelli BIM (Building Information Modeling) di edifici ad elevato valore storico. Questa stessa rappresentazione ha avuto risonanza in eventi e quaderni del CNEL (Consiglio Nazionale dell'Economia del Lavoro della Repubblica Italiana), menzione allo Smart City Expo World Congress 2022, nonché vari Keynote, Award a conferenze internazionali, e interventi su vari Master dell'Università di Firenze e negli Smart Talk dell'Università di Bologna. La modellazione Digital Twin di Snap4City ha riscosso un ampio successo, non solo nella modellazione 3D ma nella rappresentazione di elementi real time e controparti digitali complesse che sottendono al modello anche in ambito industriale.