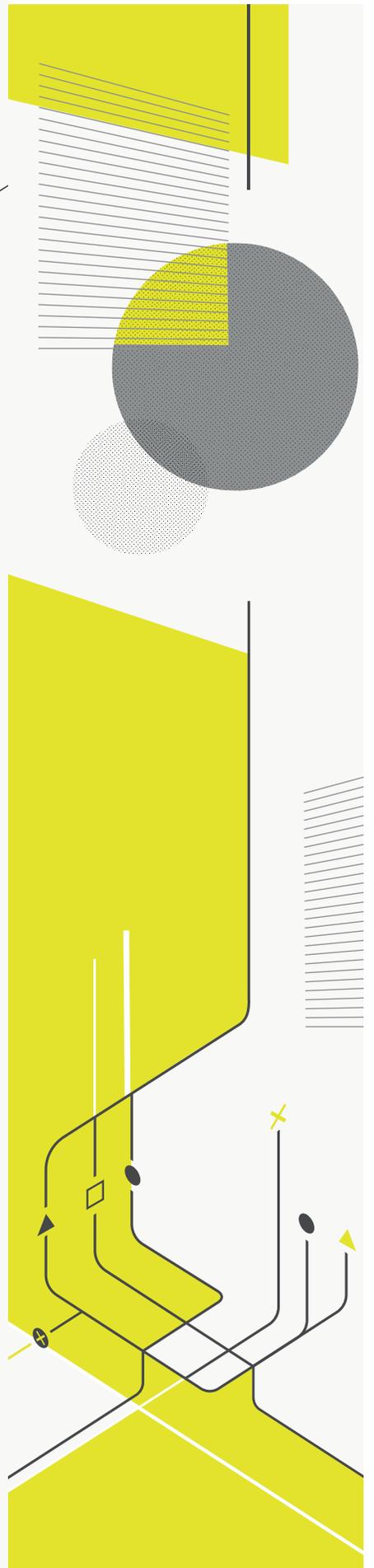


CONFERENCE  
GARR  
2019

# CONNECTING THE FUTURE

SELECTED  
PAPERS



CONFERENCE  
GARR  
2019

# CONNECTING THE FUTURE

TORINO, 4-6 GIUGNO 2019

SELECTED  
PAPERS

 Consortium  
GARR

Conferenza GARR 2019 - Connecting the future - Selected Papers

Torino, 4-6 giugno 2019

ISBN 978-88-905077-9-3

DOI 10.26314/GARR-Conf19-proceedings

Quest'opera è distribuita con Licenza Creative Commons Attribuzione 4.0 Internazionale (CC-BY).



Editore: Associazione Consortium GARR

Via dei Tizii, 6, 00185 Roma, Italia

[www.garr.it](http://www.garr.it)

Curatori editoriali: Marta Mieli, Carlo Volpe

Progetto grafico: Carlo Volpe

Impaginazione: Marta Mieli, Carlo Volpe

Prima stampa: Dicembre 2019

Numero di copie: 600

Stampa: Tipografia Graffietti Stampati snc

S.S. Umbro Casentino Km 4.500, 00127 Montefiascone (Viterbo)

Tutti i materiali relativi alla Conferenza GARR 2019 sono disponibili all'indirizzo:

[www.garr.it/conf19](http://www.garr.it/conf19)

# Indice

- 6 **Analysing Knowledge Domains that Emerge from Linked Open Data**  
Luigi Asprino, Paolo Ciancarini, Valentina Presutti
- 10 **STEM Training: preparing teachers to integrate technology and problem solving in the curriculum**  
Alice Barana, Stefania Barsottini, Cecilia Fissore, Marina Marchisio
- 14 **Ludoteca del Registro .it: un progetto di educazione digitale**  
Giorgia Bassi, Beatrice Lami
- 19 **Un ecosistema digitale per la salute della donna e del bambino: sperimentazione di una piattaforma per l'innovazione gestionale per l'IRCCS materno-infantile Burlo Garofolo**  
Michele Bava, Valeria Mallozzi, Margherita Dal Cin, Michela Cinello, Sara Zanchiello
- 24 **Social4School: educare alla consapevolezza nei social network**  
Livio Bioglio, Sara Capecchi, Valentina Di Noi, Gian Manuel Marino, Ruggero G. Pensa, Giulia Venturini
- 29 **Nuove forme di fruizione del digitale. Il repository Byterfly: la conservazione e il riuso con l'open source, access e data**  
Giancarlo Birello, Anna Perin
- 34 **Adalitix: a Web Platform for Agricultural Data Analytics**  
Luca Coviello, Marco Cristoforetti, Cesare Furlanello
- 38 **How to leverage Measurement Lab Internet metrics to get useful insights through Data Science**  
Stefania Delprete, Gianni Spalluto, Christian Racca
- 44 **TDM: un sistema aperto per l'acquisizione di dati, l'analisi e la simulazione su scala metropolitana**  
Fabio Bettio, Giovanni Busonera, Marco Cogoni, Roberto Deidda, Mauro Del Rio, Massimo Gaggero, Enrico Gobbetti, Simone Leo, Simone Manca, Marino Marrocu, Luca Massidda, Fabio Marton, Marco Enrico Piras, Luca Pireddu, Gabriella Pusceddu, Alessandro Seoni, Gianluigi Zanetti
- 50 **L'entrepreneurship education ed il progetto One Stop Shop**  
Gianluigi De Gennaro, Annamaria Demarinis Loiotile, Annalisa Turi
- 54 **Un caso riuscito di contaminazione scientifica tra ambienti diversi**  
Giuseppe Di Persio
- 59 **Managing research data at the University level: the experience of the University of Milan**  
Paola Galimberti, Susanna Mornati

- 64 **Strategia e politica di sviluppo per una infrastruttura digitale per la gestione e valorizzazione del patrimonio archivistico**  
Costantino Landino
- 69 **Costruzione di un vocabolario controllato per la terminologia sulla Cybersecurity**  
Claudia Lanza
- 73 **Una soluzione basata sulla tecnologia blockchain per la validazione di attestati e titoli rilasciati tramite piattaforma di e-learning**  
Nicola Cannistrà, Fabio Cordaro, Fabrizio La Rosa, Francesco La Rosa, Roberta Maisano, Umberto Ruggeri, Salvatore Todaro, Riccardo Uccello
- 78 **Quanti petabyte occorrono per vedere un buco nero?**  
Elisabetta Liuzzo, Kazi Rygl
- 84 **Dalla realtà proxy al panico del dasein totale. Il teatro all'epoca dell'AI e dell'AR**  
Anna Maria Monteverdi
- 89 **Connecting document interaction to support music distance learning environments: development of a multipurpose, collaborative PDF reader for real-time sharing of navigation and annotation on sheet music**  
Enrico Pietrocola
- 93 **MineHEP: Data Mining in High Energy Physics**  
Laura Redapi, Andrea Cioni, Maria Vittoria Garzelli, Andrea Ceccarelli, Piergiulio Lenzi, Vitaliano Ciulli
- 98 **MeDB: a Portable Clinical Historical Database**  
Paolo Ricciuti, Alba Di Pardo, Daniele Cafolla
- 102 **Banche dati di documenti giuridici per la formazione e le digital humanities**  
Francesco Romano, Antonio Cammelli
- 106 **Start@unito underground map for an e-learning trip**  
Marina Marchisio, Sergio Rabellino, Matteo Sacchet, Daniela Salusso
- 111 **Scuole connesse e didattica digitale. La scuola del futuro e il progetto Riconessioni**  
Giovanni Luca Spoto, Flavio Renga, Marcello Enea Newman
- 115 **Ecosistema Digitale E015: condividere l'accesso regolamentato ai dati per costruire relazioni**  
Emiliano Sergio Verga, Maurilio Zuccalà
- 120 **Il catalogo che cresce: gli autori della Beic Digital Library in Wikidata**  
Federica Viazzi
- 124 **European Open Science Cloud: infrastrutture e servizi a supporto della ricerca all'Università degli Studi di Padova**  
Lorisa Andreoli, Elena Bianchi, Yuri Carrer, Vera Dean, Antonella Zane, Michela Zorzi
- 131 **Blockchain: aspetti giuridici tra dati, numeri e algoritmi**  
Fernanda Faini
- 136 **Un authority file per i luoghi e le istituzioni: l'Atlante Storico Istituzionale dell'Italia unita**  
Stella Di Fazio



### Comitato di programma

Giuseppe Attardi - Università di Pisa - GARR  
 Claudia Battista - GARR  
 Davide Calonico - INRM e TOP-IX  
 Massimo Carboni - GARR  
 Giuseppe Corsaro - IC Leonardo da Vinci di Mascalucia e Insegnanti 2.0  
 Paolo Giaccone - Politecnico di Torino  
 Gaetano Guglielmi - Ministero della Salute  
 Leonardo Lanzi - GARR  
 Marcello Maggiora - Compagnia di San Paolo e CNR  
 Marta Mieli - GARR  
 Luisa Minghetti - Istituto Superiore di Sanità  
 Pier Luca Montessoro - Università di Udine  
 Lucia Negrin - Liceo Galilei di Trieste  
 Marisol Occioni - Università Ca' Foscari Venezia  
 Gabriella Paolini - GARR  
 Giorgio Rossi - ESFRI, Università di Milano  
 Federico Ruggieri - GARR  
 Mirella Serlorenzi - Museo Nazionale Romano  
 Sabrina Tomassini - GARR  
 Davide Vagheti - GARR  
 Stefano Vitali - Istituto Centrale per gli Archivi  
 Carlo Volpe - GARR

enti fondatori Consortium GARR  
 ENA ENET INFN

Consortium GARR THE ITALIAN EDUCATION & RESEARCH NETWORK

Conferenza GARR 2019  
**CONNECTING THE FUTURE**  
 Torino 4-6 giugno 2019

ospitato da  
 POLITECNICO DI TORINO  
 Aula Magna  
 corso Duca degli Abruzzi 24

main sponsor  
 HUAWEI

sponsor  
 TIM  
 FORTINET

Tutte le presentazioni e maggiori informazioni sono disponibili sul sito dell'evento:  
[www.garr.it/conf19](http://www.garr.it/conf19)

# Analysing Knowledge Domains that Emerge from Linked Open Data

Luigi Asprino<sup>1,2</sup>, Paolo Ciancarini<sup>2</sup>, Valentina Presutti<sup>1,2</sup>

<sup>1</sup>STLab, ISTC-CNR, Roma <sup>2</sup>Università di Bologna

**Abstract.** The aim of the Semantic Web initiative is to create a Web in which data is represented through symbols with a shared syntax and formal semantics. Several analyses have sought to investigate how these formal languages are used in practice, but very few of them analysed the Semantic Web per knowledge domain. In this paper we present a novel approach for analysing Semantic per knowledge domain in bottom-up fashion leveraging on topic modeling and natural language processing techniques.

**Keywords.** Semantic Web, Linked Open Data, Empirical Semantics

## Introduction

The aim of the Semantic Web initiative is to create a Web in which data is represented through symbols with a shared syntax and semantics (Berners-Lee et al., 2001). This vision enacts machines to autonomously exchanging, analysing and using data found on the Web for their tasks, thus making the Web a huge knowledge base for intelligent agents. In the last 20 years the research, industry, and public administration communities have contributed to make real this vision by giving birth to the Linked Open Data: a huge network of ~200 billions<sup>1</sup> linked facts formally and uniformly represented in RDF and OWL.

The collection of Linked Open Data (LOD) datasets forms the largest publicly accessible Knowledge Graph (KG) that is available on the Web today. Over the years, many studies have analysed these datasets, often focusing on the structure and dimension of the data, as well as providing statistics that shed light on the composition of the datasets. In most cases, such observations have been based on relatively small samples of the LOD KG. Moreover, it is not always clear how representative the chosen samples are. This is especially the case when observations are based on one dataset (e.g., DBpedia), or a handful of datasets, draw from the much larger LOD KG.

Linked Open Data and Semantic Web ontologies are encoded using RDF facts and/or OWL axioms, thereby exhibiting a formal semantics. Several analyses have sought to investigate how these formal languages are used in practice: how are certain formal constructs (e.g., owl:sameAs identity) used, and to what extent are LOD guidelines (e.g., limiting the use of blank nodes) followed in practice? Very few of them analysed LOD per knowledge domain, namely: how are certain formal constructs used in linguistic or government knowledge domain? To what extent are LOD guidelines followed in practice in encyclopedic or geographical knowledge domain? Most of proposed approaches for performing this kind of analyses rely on metadata provided by `\url{lod-cloud.net}` that

<sup>1</sup>This estimation is obtained from metadata provided by [www.lod-cloud.net](http://www.lod-cloud.net)

specifies provenance and knowledge domain of LOD datasets (e.g. Schmachtenberg et al., 2014). However, most of LOD datasets are not provided with suitable metadata for such an analysis:

- The vocabulary used for specifying the knowledge domain of LOD datasets is poor (it consists of 9 labels: Cross-domain, Geography, Government, Life Sciences, Linguistics, Media, Publications, Social Networking, User-Generated).
- ~25% of LOD datasets do not indicate their knowledge domain, while the others indicate a single label for whole datasets.

Moreover, most of these labels are given in a top-down fashion with the risk that label and dataset may result partly uncorrelated.

This paper presents a novel approach for analysing LOD per knowledge domain that relies on topic modeling and natural language processing techniques. Topic modeling is a text-mining technique for discovering the topics that occur in a collection of documents. Intuitively, given a document about a particular topic (e.g. "Religion"), one would expect words related to that topic to appear in the document more or less frequently than others (e.g. "God" will appear more often than "Pizza" in document about religion). The "topics" produced by topic modeling techniques are lists (one for each topic) of weighted words. Specifically, a topic  $t$  is a list of pairs  $(w_i, p_i)$  where  $w_i$  is a word and  $p_i$  is the likelihood that  $w_i$  appears in a document about the topic  $t$ . A document typically concerns multiple topics in different proportions: for example in a document that is 10% about science and 90% about religion, there would probably be about 9 times more words about religion than those about science.

## 1. Proposed Approach

The collection of documents to analyse with topic modeling techniques is created as follows. For each LOD dataset  $d$ , we create a virtual document by concatenating natural language descriptions associated with entities within  $d$ . A LOD entity can be associated with two natural language descriptions:

- a label, a short text content used for naming the entity which is indicated by the property `rdfs:label`;
- a comment, a description of the resource in natural language, often providing examples of the concept being defined which is indicated the property `rdfs:comment`.

Then, a topic modeling library is used in order to extract the topics that emerge from virtual documents. The emerging topics will be inspected and manually aligned with a well-known taxonomy of knowledge domains (e.g. WordNet's taxonomy). Finally, the extracted topic model will be used to associate virtual documents (hence, LOD datasets) with the extracted topics. As a result, LOD datasets will be annotated (in a bottom-up fashion) with topics aligned to a well-known taxonomy of knowledge domains, thus enabling an analysis of LOD datasets per knowledge domain.

## 2. Current status of the work

As input dataset for our analysis we used the crawl provided by LOD Laundromat (Beek

et al., 2014) (a project aimed at crawling data dumps published as part of the LOD cloud). For each dataset crawled by LOD Laundromat, we computed a virtual document. This process took 11.5 hours for analysing 640M triples on a m1.xxl instance (64 GB RAM, 16 virtual CPUs) provided by the GARR's cloud platform. The resulting dataset is provided in TSV format and available at the following link . In the next months the dataset will be analysed with the latent semantic analysis (LSA)\footnote{We are using the implementation of LSA provided by Gensim . LSA is a topic modelling technique that assumes that words with related meaning will occur in similar context. A word-document matrix containing word counts per document is constructed from the dataset (rows of the matrix represent unique words and columns represent documents). Then, a mathematical technique called singular value decomposition (SVD) is used to reduce the number of rows while preserving the similarity structure among columns. The resulting rows with their associated scores for each document will constitute the extracted topics.

## References

Tim Berners-Lee, James Hendler, and Ora Lassila. (2001), The Semantic Web, *Scientific american* (284.5), pp 34–43.

Max Schmachtenberg, Christian Bizer, and Heiko Paulheim. (2014) Adoption of the Linked Data Best Practices in Different Topical Domains, *Proceedings of the 13th International Semantic Web Conference (ISWC 2014), Part I*, pp 245– 260.

Wouter Beek, Laurens Rietveld, Hamid R. Bazoobandi, Jan Wielemaker, and Stefan Schlobach. (2014), LOD Laundromat: A Uniform Way of Publishing Other People's Dirty Data, *Proceedings of the 13th International Semantic Web Conference (ISWC 2014), Part I*, pp 213–228.

<https://wiki.dbpedia.org>

<http://lodlaundromat.org>

<https://w3id.org/edwin/garr2019>

<https://radimrehurek.com/gensim>

## Authors



**Luigi Asprino** - [luigi.asprino@istc.cnr.it](mailto:luigi.asprino@istc.cnr.it)

Luigi Asprino is a research assistant at the Institute of Cognitive Science and Technologies of National Research Council in Italy. He received a PhD in Computer Science and Engineering in 2019 from the University of Bologna. He has worked in national and european projects: MARIO (EU), MARE (EU) ArCo (IT), EcoDigit (IT). He has been involved in the organisation ESWC 2017 and he has served as program committee member and reviewer for many international conferences such as: ESWC, ISWC, WOP, EKAW, IJCAI and AAAI. His research interests include Knowledge Representation and Reasoning, Natural Language Processing and Machine Learning.

**Paolo Ciancarini** - [paolo.ciancarini@unibo.it](mailto:paolo.ciancarini@unibo.it)

Paolo Ciancarini is Professor of Computer Science at the Univ. of Bologna. He got a Phd in Informatics at the University of Pisa.

In Bologna he lectures on Software Engineering and Software Architecture, and is member of the Faculty of the PhD School in Computer Science. He currently is the President of the Italian Association of University Professors in Computer Science. He is the author of over 200 scientific papers and books. He is married, has two children, and is a passionate chess player and book collector.



**Valentina Presutti** - [valentina.presutti@cnr.it](mailto:valentina.presutti@cnr.it)

Valentina Presutti coordinates the Semantic Technology Laboratory of the National Research Council (CNR) in Rome. She received her Ph.D in Computer Science in 2006 at University of Bologna (Italy). She has coordinated, and worked as researcher in, many national and european projects on behalf of CNR – some examples: IKS (EU), MARIO (EU), NeOn (EU), ArCo (IT), EcoDigit (IT). She co-directs the International Semantic Web Research Summer School (ISWS). She serves in the editorial board of top journals such as Journal of Web Semantics, Journal of the Association for Information Science and Technology, Data Intelligence Journal, Intelligenza Artificiale. And she has been involved in the organisation of top semantic web conferences such as ISWC and ESWC and she is senior PC for IJCAI. She is one of the creators of the ontologydesignpatterns.org initiative and of the workshop series on Ontology Design and Patterns (WOP). She has 100+ publications in international journals/conferences/workshops on topics such as semantic web, knowledge extraction, and ontology design. She teaches Social Robotics and Programming as adjunct professor at the University of Bologna, and collaborates as scientific and technological expert for private as well as public organizations. Her research interests stand at the crossing between semantic web and artificial intelligence and include knowledge graphs, empirical analysis of the semantic web, social robotics.

# STEM Training: preparing teachers to integrate technology and problem solving in the curriculum

Alice Barana<sup>1</sup>, Stefania Barsottini<sup>2</sup>, Cecilia Fissore<sup>1</sup>, Marina Marchisio<sup>1</sup>

<sup>1</sup>Università di Torino, <sup>2</sup>Liceo Scientifico Statale "Galileo Ferraris" di Torino

**Abstract.** This paper presents and discusses the design and experimentation of a teacher training course on a set of digital and methodological tools aimed at enhancing the teaching and learning of STEM in a problem solving perspective at secondary school level. The course included face-to-face lessons, online meetings, assignment submissions, group-work, laboratories, and the application of the methodologies learned in the classroom. It was attended by 39 teachers in the 2017/2018 school year. Data about appreciation and effectiveness of the training modalities and the perceived usefulness of the proposed methodologies for learning, collected through questionnaires, are analyzed and discussed, with the purpose of gaining insights into how to prepare teachers to integrate technology and problem solving in the curriculum.

**Keywords.** Advanced Computing Environment, Automatic Assessment, Problem Posing and Solving, STEM Education, Teacher Training, Digital Learning Environment.

## 1. Introduction and theoretical framework

In STEM (Science, Technology, Engineering, Mathematics) education, even at secondary level, an approach based on interdisciplinary problem solving is highly recommended: it consists of using real-world relevant problems as core units, around which meaningful contents can be developed. This teaching method facilitates the development of disciplinary competences and soft skills, such as teamwork, self-regulation and problem-solving attitude, which are more and more important for employment. Digital technologies enable students to tackle relevant problems, as they can support data analysis, modeling and exploration (Carreira et al., 2016).

Our research group, who works on digital education at the University of Turin, proposes a set of technological and methodological tools aimed at enhancing the teaching and learning of STEM in a problem solving perspective:

- the use of an Advanced Computing Environment (ACE) for problem solving. An ACE is a system able to perform numeric and symbolic computation, geometric visualization in 2 and 3 dimensions, embedding of interactive components and programming (Barana et al., 2017a);
- formative assessment with an Automatic Assessment System (AAS). In particular, we propose the use of contextualized tasks with an AAS, whose questions are based on algorithms which support the creation of formulas, graphs, random parameters and open mathematical answers automatically assessed for their equivalence to the correct solution. Immediate and interactive feedback elicit the development of problem solving

strategies (Barana et al., 2018);

- the integration of interactive materials into a Digital Learning Environment (DLE), where students can practice teamwork, self and peer assessment in a learning community. Italian teachers, whose average age is among the oldest in Europe, are not used to encompassing these methodologies in teaching: they need to be properly trained in order to efficiently use innovative technologies in their daily practice. One of the most widespread models for training teachers to integrate technology into teaching is the TPACK (Technological, Pedagogical And Content Knowledge) model. It suggests to assemble three kinds of knowledge in teacher training: the technical use of the tools, the pedagogical aspects related to learning with technologies, and the disciplinary sphere of knowledge (Voogt & McKenney, 2017). We believe that a deep knowledge is certainly necessary, but not sufficient in order to prepare teachers to actually integrate the technology in their lessons. They need support when putting into practice what they learned, and reflection on practice in order to analyze their actions and improve their methods. They need to feel assisted, by experts and by colleagues, they need to fill part of a community they can rely on.

## 2. Methodology

Bearing this framework in mind, we designed a teacher training course for STEM secondary school teachers, called “STEM TRAINING”, on the use of the technological and methodological set of tools above mentioned in interdisciplinary problem solving activities. The training course was shaped according to the model of teacher training developed by our research group (Barana et al., 2017b). It includes face-to-face lessons, online meetings, assignment submissions, group-work, laboratories, and the application of the methodologies learned in the classroom. The structure of the course is the following:

- 5 face-to-face weekly meetings lasting 3 hours each, held by trainees who introduced teachers to problem posing, problem solving, the creation of a digital environment for students, the use of an ACE for problem solving, formative assessment with an AAS. During the meetings, teachers could work, individually or in groups, towards the creation of didactic materials. After each meeting, teachers had to submit a didactic material related to the topic of the meeting, to be used with their students;
- a virtual community of practice was created in an e-learning platform where teachers could collaborate in the creation of digital materials, have access to the training materials and ask for the tutors’ support;
- 10 weekly online meetings, lasting one hour each, aimed at deepening the topics dealt with in the course. The meetings were held through a web-conference tool which allowed tutors to share their desktop and teachers to interact through voice and chat;
- 10 afternoon laboratories, lasting 3 hours each, on specific disciplinary or interdisciplinary topics, such as Physics and Science experiments, programming languages, data base management systems or virtual simulation labs;
- experimentation of the materials produced according to the methodologies presented with the teachers’ classes.

STEM TRAINING was activated in the 2017/2018 school year, from November to May. It was promoted by ReLiSPi, the scientific high schools network of Piedmont, but it was open to teachers of all lower and upper secondary schools.

Teachers could attend the different parts of the course according to their interests; at the end of the course they received a certification testifying the number of hours dedicated to the training. Time spent in asynchronous work was taken into account as well: each submission of didactic materials (those required after the face-to-face meetings) accounted for 3 training hours, while the experimentation of each activity with the class accounted for 5 hours.

Before and after each training part, teachers were asked to fill in brief questionnaires, inquiring about the appreciation and effectiveness of the training modalities and the perceived usefulness of the proposed methodologies for learning. There was room for reflecting about classroom episodes, both through the questionnaires' items and in the online forums, where teachers could share their experiences and discuss them.

### **3. Results and conclusions**

39 teachers enrolled in the course; 5 of them taught Mathematics and Science at lower secondary schools while the others taught different STEM subjects at upper secondary schools, including Mathematics, Physics, Science and Technical disciplines; moreover, 2 Philosophy teachers joined the course, as they were interested in the scientific approach to problem solving and in interdisciplinary issues.

Most of the teachers regularly attended the course's activities; the average number of training hours certified at the end of the course was 37.16 with a standard deviation of 18.20, ranging from 6 (a couple of teachers only attended 2 face-to-face meetings) to 67 (the most assiduous attenders). Data from the questionnaires show that they largely appreciated the training modalities (they evaluated them with 4 points out of 5 in a Likert-scale), in particular working with teachers of other disciplines and the possibility to create and share teaching materials. After experimenting the innovative methodologies with their students, the teachers assessed in a Likert scale from 1 to 5 the effectiveness of these kinds of activities in relation to several didactic purposes. It resulted that the DLE was considered useful to propose problems and exercises (mean: 4.2), to increase students' motivation (4.0) and their interest in studying (3.8). The ACE was considered effective especially to better understand the contents (3.7) and to develop competence (3.8), while the AAS to provide students with personalized feedback (4.3) and increase their awareness of their abilities (4.2). Problem posing and solving was appreciated for helping students develop reasoning and abstraction capabilities (4.5), critical thinking (4.3) and teamwork skills (4.3).

### **References**

Barana, A., Conte, A., Fioravera, M., Marchisio, M., & Rabellino, S. (2018), A Model of Formative Automatic Assessment and Interactive Feedback for STEM, Proceedings of 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), 1016–1025.

Barana, A., Fioravera, M., & Marchisio, M. (2017a), Developing problem solving competences through the resolution of contextualized problems with an Advanced Computing Environment, Proceedings of the 3rd International Conference on Higher Education Advances, 1015–1023.

Barana, A., Fioravera, M., & Marchisio, M. (2017b), Teacher training: A model for introducing innovative digital methodologies for learning Mathematics, Proceedings of the 3rd International Conference on Higher Education Advances, 608–616.

Carreira, S., Jones, K., Amado, N., Jacinto, H., & Nobre, S. (2016), Youngsters Solving Mathematical Problems with Technology, Springer.

Voogt, J., & McKenney, S. (2017), TPACK in teacher education: Are we preparing teachers to use technology for early literacy?, Technology, Pedagogy and Education, 26(1), 69–83.

## Authors



Alice Barana - [alice.barana@unito.it](mailto:alice.barana@unito.it)

Alice Barana is graduated in Mathematics and a PhD student in Mathematics Education at the Department of Mathematics of the University of Turin. She is also a secondary school teacher of Mathematics. Her research deals with innovative methodologies for learning STEM subjects based on the use of digital learning environments and automatic assessment as means to improve learning and self-regulation. She joins several research projects on e-learning, digital education, school and academic success.

Stefania Barsottini - [dirigente@liceogalfer.it](mailto:dirigente@liceogalfer.it)

Stefania Barsottini is the head teacher of Liceo Scientifico Galileo Ferraris of Turin. She bases the management of her school on innovation and digitalization, promoting several activities aimed at improving disciplinary and cross-cutting competences, in collaboration with the University of Turin and other companies and institutions. She is the coordinator of ReLiSPi, the scientific high school network of Piedmont, which organizes actions of teacher training and innovative projects for students.



Cecilia Fissore - [cecilia.fissore@unito.it](mailto:cecilia.fissore@unito.it)

Cecilia Fissore is graduated in Mathematics and she is a PhD student in Digital Humanities at the University of Turin and Genoa. She collaborates on numerous research projects and she studies innovative methodologies for STEM education: Virtual Learning Environment for collaborative learning; Advanced Computing Environment for problem solving and Automated Assessment with interactive feedback. She has experience in the design and the management of teacher training courses.

Marina Marchisio - [marina.marchisio@unito.it](mailto:marina.marchisio@unito.it)

Marina Marchisio is Professor of Mathematics at the University of Turin, Department of Molecular Biotechnology and Health Sciences. Her research domain is Digital Education, in particular teaching and learning STEM disciplines with new technologies. She is the scientist-in-charge of the Italian Ministry of Education's Project Problem Posing and Solving and the coordinator of several Projects of the University of Turin for digital university guidance, e-learning, school and academic success.



# Ludoteca del Registro .it: un progetto di educazione digitale

Giorgia Bassi, Beatrice Lami

CNR IIT, Istituto di Informatica e Telematica

**Abstract.** Il contributo descrive il progetto Ludoteca del Registro .it dell'Istituto di Informatica e Telematica del CNR di Pisa, nato con l'obiettivo di diffondere tra i bambini delle scuole primarie l'utilizzo consapevole e sicuro della Rete Internet.

**Keywords.** Scuola digitale, cultura digitale, Internet, sicurezza informatica

## Introduzione

La Ludoteca del Registro .it ([www.ludotecaregistro.it](http://www.ludotecaregistro.it)), patrocinato dall'Autorità Garante per l'Infanzia e l'Adolescenza, è un progetto di educazione digitale per le scuole primarie avviato nel 2012 dal Registro .it, l'anagrafe dei domini .it (gestito dall'Istituto di Informatica e Telematica del Cnr di Pisa).

I laboratori affrontano i temi della Rete, in tutti i suoi aspetti, dalla struttura tecnica, alla storia e organizzazione, per arrivare ai temi legati alle risorse (social media, banche dati ed enciclopedie online, internet delle cose, smart city) e ad un loro utilizzo sicuro e responsabile.

Ad oggi la Ludoteca ha incontrato oltre 11.000 bambini in tutto il territorio nazionale, per un totale di 1100 ore di formazione. Il progetto si rivolge alla “generazione Z” (i nati dal 1995 ad oggi) che utilizza le risorse della Rete con notevole abilità tecnica, passando con naturalezza dallo spazio fisico a quello cyber, senza però una adeguata conoscenza dei potenziali rischi.

A questo riguardo, riportiamo alcuni dati della ricerca “EU Kids Online per MIUR e Parole O\_Stili” del 2018, dalla quale emerge che lo smartphone è il principale strumento con cui questo target (ma in generale l'intera popolazione) accede a Internet: è usato quotidianamente per andare online dal 97% dei 15-17enni e dal 51% dei bambini di 9-10 anni. Smartphone e tablet hanno quindi ampliato le coordinate spazio-temporali dell'uso di internet, fornendo un accesso “anywhere, anytime”.

Sul fronte dei rischi, i dati presentano uno scenario non molto incoraggiante: il 25% dei 15-17enni non parla con nessuno delle esperienze negative su Internet, il 27% risolve il problema chiudendo semplicemente la pagina web o l'app; il 22% di chi ha avuto un'esperienza negativa sui social dichiara di aver reagito bloccando un contatto, il 10% modificando le proprie impostazioni di privacy; e solo il 2% segnalando contenuti o contatti inappropriati ai gestori delle piattaforme.

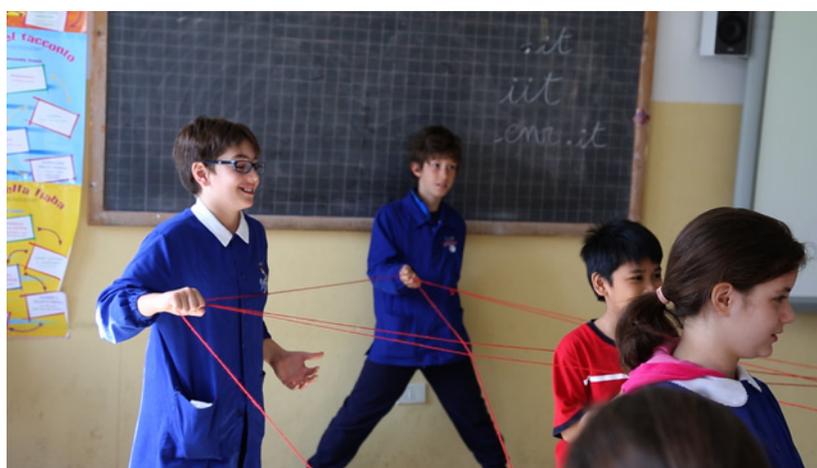
## 1. Il progetto Ludoteca del Registro. it: target, contenuti e strumenti

La Ludoteca si rivolge ai bambini delle scuole primarie con l'obiettivo di diffondere la cultura di Internet, partendo dalla conoscenza funzionale delle tecnologie digitali.

Il tema della Rete Internet è introdotto a partire dalle basi tecniche (linguaggio binario, trasmissione dei dati, nomi a dominio, indirizzo IP, protocolli), per trattare, in seguito, le principali problematiche legate all'utilizzo del web, come ad esempio la tutela della propria identità digitale o l'attendibilità dei contenuti.

Le lezioni nelle classi prevedono una parte di introduzione al tema, seguita da attività di gruppo, svolte con strumenti molto semplici (materiale cartaceo, lavagna, pennarelli), in modo da consentire la partecipazione anche a scuole meno dotate da un punto di vista tecnologico. Dal punto di vista delle metodologie didattiche si cerca di rompere lo schema della lezione frontale, stimolando il dibattito e coinvolgendo gli alunni in attività di "collaborative learning".

Fig. 1  
Il gioco del filo,  
usato per spiegare  
la trasmissione dei  
dati via Internet



## 2. Internetopoli: alla scoperta della Rete

L'esigenza di diffondere il progetto su larga scala ha portato a realizzare nel 2015 la web app "Internetopoli", compatibile con la Lim e scaricabile all'indirizzo <http://www.internetopoli.it>.

L'app offre ai docenti un percorso guidato che permette di lavorare in classe anche in totale autonomia, consultando il materiale di supporto e approfondimento a corredo dello strumento. I contenuti dell'app ripropongono i temi dei laboratori nelle classi e si articolano in otto livelli di navigazione (come funziona Internet, i nomi a dominio, storia e organizzazione, social media, privacy in rete, informazioni e contenuti online, internet delle cose, smart city).

La metafora della città lega coerentemente tutti gli argomenti, da quelli più tecnici a quelli legati alle risorse del web: l'indirizzo postale delle varie case della città di Internet diventa, per esempio, l'indirizzo IP della Rete internet, univoco e quindi strumento di rintracciabilità.

A supporto dell'app è disponibile anche la "Guida per gli insegnanti", nella quale

Fig. 2  
L'applicazione  
Internetopoli,  
screenshot



sono presenti tutti i contenuti dei livelli, riportati in un ordine lineare, con l'aggiunta di materiale di approfondimento e proposte di percorsi didattici da svolgere in classe. E' possibile inoltre scaricare il gioco stampabile "Internetopoli-The Game", una sorta di gioco dell'oca ambientato nella città di Internet.

Dal sito internetopoli.it si accede anche alla sezione "Giochi" che propone quattro videogiochi sui temi della Rete: il "Memory" (abbinamento di immagini identiche tutte relative all'ambito informatico e alla tecnologia), "Trova la coppia" (abbinamento logico di due immagini diverse, come ad esempio un mouse e una tastiera), il "Gioco dell'oca" e il "Quiz on the net" con domande a risposta multipla sui temi affrontati dall'app.

### 3. Giocare con la cyber security

Il tema della sicurezza informatica è di assoluta attualità ma non sempre si hanno le conoscenze necessarie per individuare le situazioni rischiose. A maggior ragione, diventa difficile mettere in guardia i bambini da situazioni pericolose e insegnare loro comportamenti corretti. I laboratori di cyber security affrontano alcune delle situazioni potenzialmente più rischiose della vita online: richieste di contatto da parte di sconosciuti, furto di dati personali, download da siti poco affidabili, connessioni a reti non sicure. L'obiettivo è fare cultura della sicurezza, stimolando un approccio preventivo, attraverso la conoscenza dei rischi e dei comportamenti da adottare.

Fig. 3  
I laboratori di  
cyber security,  
tavole a  
fumetti



I lab sono introdotti da una breve presentazione che, attraverso immagini di vita quotidiana, evidenzia l'importanza di conoscere le risorse da tutelare e le potenziali minacce.

Parte del materiale cartaceo utilizzato per i giochi consiste in vignette a fumetti, ambientate ad Internetopoli e con protagonista Nabbovaldo (unione di Nabbo che nel gergo dei millennials indica un "novellino" e il personaggio Marcolvaldo di Calvino), un ragazzo che usa Internet e le tecnologie digitali senza essere consapevole dei potenziali rischi. Attraverso il gioco del "Cifrario di Giulio Cesare" si introduce anche il tema della crittografia, una delle tecniche più usate in Rete per assicurare la riservatezza dei messaggi.

#### 4. Peer education ed educazione digitale

Per venire incontro alle crescenti richieste di partecipazione in tutto il territorio nazionale, è stato necessario rendere scalabile il progetto, in un primo tempo grazie alla web app Internetopoli e, in seguito, introducendo il modello della "peer education".



Fig. 4  
Il progetto  
Let's Bit!

È nato quindi il progetto Let's Bit!, gestita attraverso l'alternanza scuola/lavoro, al quale partecipano i ragazzi degli istituti superiori. Dopo un'adeguata formazione, gli studenti diventano educatori della Rete nelle classi primarie, ruolo che li responsabilizza e che li aiuta a riflettere con maggiore attenzione sui propri comportamenti online. Ad oggi hanno partecipato al progetto su tutto il territorio nazionale e non solo (ha partecipato anche un istituto superiore bilingue in Albania):

- 540 ragazzi
- 35 classi
- 180 ore di formazione

#### 5. Conclusioni

Il progetto Ludoteca del Registro .it rappresenta oggi un unicum nel panorama delle iniziative educative dedicate alla Rete, per il tipo di destinatario (le scuole primarie) e per il tipo di contenuti affrontati, che spaziano dal funzionamento della Rete, alla cyber security, per arrivare alle nuove frontiere dell'Internet delle cose.

Tra gli obiettivi a breve e lungo termine sicuramente la scalabilità, ovvero rendere il progetto fruibile in tutto il territorio nazionale, attraverso strumenti digitali e modelli didattici alternativi, come ad esempio la peer education, già adottata con risultati molto positivi per il progetto Let's Bit!

## Riferimenti bibliografici

Ferri P., (2008), La scuola digitale, Mondadori, Milano.

Calvani A., (2013), I nuovi media nella scuola. Perché, come, quando avvalersene, Carocci, Roma.

AA. VV, (2009), Didattica attiva con la Lim, Erickson, Trento.

Maglioni M., Biscaro F., (2014), La classe capovolta. Innovare la didattica con la flipped classroom, Erikson, Trento.

Pellai A., Rinaldin V., Tamborini B., (2002), Educazione tra pari. Manuale teorico-pratico di empowered peer education, Erikson, Trento.

## Autrici



**Giorgia Bassi** - [giorgia.bassi@iit.cnr.it](mailto:giorgia.bassi@iit.cnr.it)

Master in Comunicazione e Multimedia, dal 2006 lavora all'Istituto di Informatica e Telematica del Cnr di Pisa in cui ha sede il Registro .it l'anagrafe dei nomi a dominio a targa .it, collaborando a progetti di comunicazione legati ai nomi a dominio. Dal 2011 cura i contenuti, la comunicazione e le attività di divulgazione della Ludoteca del Registro .it. È referente del progetto di peer education Let's Bit! destinato agli istituti superiori.

**Beatrice Lami** - [beatrice.lami@iit.cnr.it](mailto:beatrice.lami@iit.cnr.it)

Laurea Magistrale in Informatica, Master in Management della Formazione. Dal 2000 lavora all'Istituto di Informatica e Telematica del Cnr di Pisa in cui ha sede il Registro .it. Si occupa di aspetti tecnici legati alla registrazione dei nomi a dominio, della formazione dedicata ai Registrar; dal 2011 è referente del progetto Ludoteca del Registro .it, di cui valida anche i contenuti tecnici. Collabora anche al progetto di peer education Let's Bit!



# Un ecosistema digitale per la salute della donna e del bambino: sperimentazione di una piattaforma per l'innovazione gestionale per l'IRCCS materno-infantile Burlo Garofolo

Michele Bava<sup>1</sup>, Valeria Mallozzi<sup>1,2</sup>, Margherita Dal Cin<sup>3</sup>, Michela Cinello<sup>4</sup>, Sara Zanchiello<sup>4</sup>

<sup>1</sup>IRCCS materno-infantile Burlo Garofolo, Ufficio Sistema Informativo – SC Ingegneria Clinica, Informatica e Approvvigionamenti, <sup>2</sup>Università degli Studi di Trieste, DIA – Dipartimento di Ingegneria e Architettura, <sup>3</sup>IRCCS materno-infantile Burlo Garofolo, Direzione Strategica, <sup>4</sup>Area Science Park, Ufficio Innovazione e Complessità

**Abstract.** In Sanità, sta emergendo sempre più la necessità di utilizzare sistemi tecnologici che consentano lo scambio di dati fra le strutture della rete sanitaria. Questa necessità, unita al crescente utilizzo da parte del cittadino di strumenti informatici, come app a contenuto sanitario, senza nessuna garanzia di qualità e privacy, ha portato a un progetto congiunto tra l'IRCCS "Burlo Garofolo" e Area Science Park. L'obiettivo di questo studio è quello di attivare un'azione pilota nell'ambito dei processi socio-sanitari, con un approccio volto all'innovazione gestionale. Si vuole realizzare una piattaforma sulla quale far convergere diversi servizi, incluse due app, una riguardante il percorso nascita e una rivolta al paziente diabetico. Tra le funzioni delle app, è prevista la pianificazione di attività, la comunicazione protetta e sicura con i professionisti, l'inserimento di misure cliniche e di laboratorio; inoltre sono previste sezioni Q&A e un chatbot a supporto di tutti i processi.

**Keywords.** App a contenuto sanitario, piattaforma digitale, innovazione gestionale, chatbot.

## Introduzione

I sistemi ICT in uso presso la Regione Friuli Venezia-Giulia includono servizi a sostegno dei processi ospedalieri interni e di scambio ed interazione con i cittadini. Attualmente le soluzioni offerte sono per attività di carattere amministrativo, quali la prenotazione, la disdetta, il pagamento di prestazioni sanitarie, lo scarico di referti online, accessibili mediante sistemi gestionali clinici e mediante i portali sanitari della Regione. Questi sistemi in Sanità sono ampiamente sottoutilizzati rispetto ad altri ambiti ma, parallelamente, si riscontra un uso crescente di strumenti informatici non sicuri, sia tra professionisti che tra professionisti e cittadini (ad es. teleconsulto via mail, whatsapp e facebook), senza alcuna regolamentazione, tutela della privacy e contabilizzazione dell'attività. Oltre a questo aspetto, bisogna considerare anche una crescente autogestione del cittadino, il quale ricorre alla ricerca online, alla condivisione di informazione sui social media, ad app a contenuto sanitario o legate all'area "wellness/fitness", quindi fonti non certificate, da

cui si ricavano informazioni non sempre attendibili e potenzialmente pericolose. A partire da queste considerazioni, l'obiettivo di tale progetto è quello di perseguire il miglioramento della qualità, della sicurezza e dell'appropriatezza degli interventi sanitari anche attraverso sistemi ICT, che consentano lo scambio di informazioni clinico-diagnostiche fra le strutture della rete sanitaria, la raccolta delle informazioni critiche per il controllo di gestione, il monitoraggio dei percorsi e la valutazione degli esiti, nonché lo svolgimento delle attività di ricerca.

Si è scelto di implementare tale progetto presso l'IRCCS "Burlo Garofolo" di Trieste, il quale risulta essere il contesto ideale per dimensioni contenute e tipologia di pubblico afferente; inoltre, è identificato come presidio ospedaliero specializzato nell'area materno-infantile, come "Hub" di secondo livello per le funzioni dell'area pediatrica, ed è parte integrante della rete Pediatrica Regionale. Lo scopo è quello di garantire lo stesso, elevato, livello di salute a tutti i bambini della Regione, a prescindere dalla sede di erogazione delle cure, concentrando le competenze, e garantendo una "rete di prossimità" per i bisogni più semplici e frequenti, in modo da spostare il bambino solo quando è strettamente necessario, e quando possibile far spostare i professionisti o le informazioni. La sperimentazione, qualora i risultati siano positivi e funzionali, potrà essere estesa alle altre Aziende del SSR.

## 1. Descrizione del progetto: obiettivi, servizi e strumenti

Il progetto, come schematizzato in Fig. 1, consiste nella creazione di una piattaforma sulla quale far convergere i servizi di eccellenza per: potenziare l'empowerment del cittadino nei confronti dell'offerta sanitaria; migliorare gli scambi di conoscenza tra i professionisti; agevolare la consapevolezza, condivisione, conoscenza biomedica a favore dei processi di cura ma anche di previsione; effettuare un'analisi epidemiologica e statistica; perseguire il raggiungimento di obiettivi funzionali (maggior efficienza), emozionali (il paziente e la famiglia devono sentirsi sicuri) e sociali; proporre un nuovo modello di gestione in ambito socio-sanitario; convalidare un modello di sostenibilità e collaborazione tra enti pubblici e privati. I servizi che verranno implementati sono: teleconsulto tra professionisti, interazione medico-utente, informazione/formazione del cittadino, supporto alla cura e al monitoraggio.

L'intera piattaforma (certificata medicale 93/42/CE in classe IIA) sarà integrata in un da-

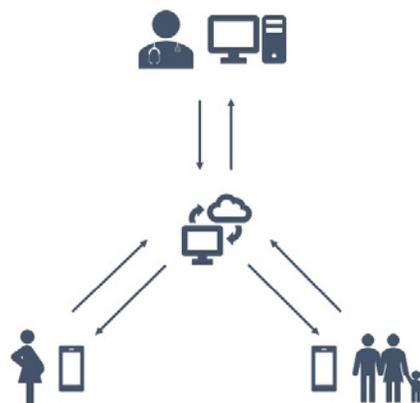


Fig. 1  
Schema del progetto

ta center ISO 27001 e dotata di un Call Center di supporto agli operatori e ai pazienti. L'architettura della piattaforma (online, interattiva e personalizzata) sarà modulare e capace di integrare una serie di strumenti (siti/portali web) che, condividendo la stessa knowledge base, possano dare supporto ai diversi servizi e applicativi, come rappresentato schematicamente in Fig. 2. In particolare, sono in fase di sviluppo rispettivamente un'app che implementi un servizio di presa in carico della gestante e dei neonati fino ai primi anni di vita, ed una per la gestione della cronicità (diabete) dei pazienti pediatrici. Tali applicazioni permettono la pianificazione di attività, appuntamenti, e della compliance; la comunicazione protetta e sicura con i professionisti; l'inserimento di misure cliniche; la pubblicazione da parte dei professionisti di documentazione multimediale e video; la creazione di una sezione Q&A e di un chatbot a supporto dei processi, con una specificità per aspetti sanitari e clinici.

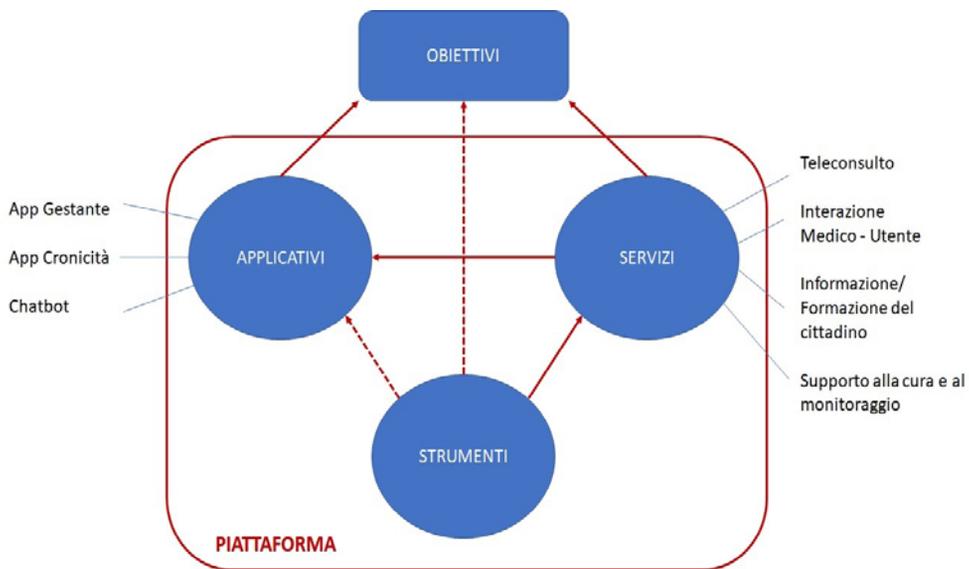


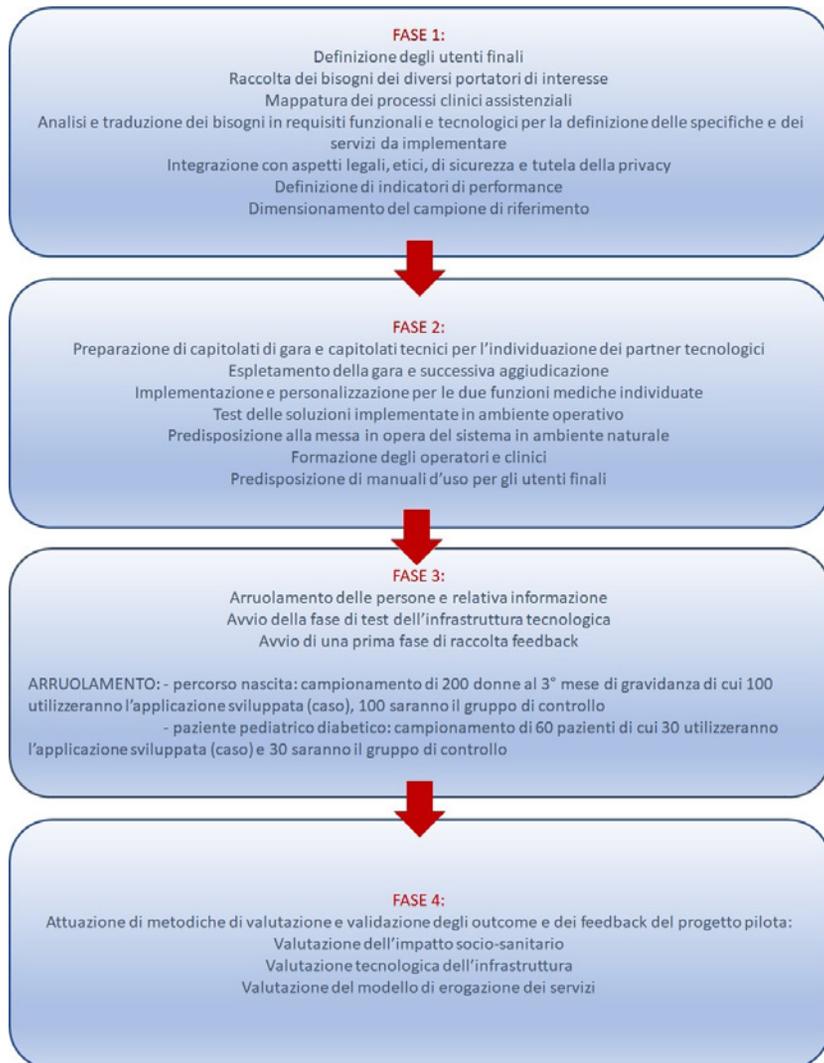
Fig. 2  
Architettura della piattaforma

## 2. Le fasi del progetto

Le fasi del progetto, descritte in Fig. 3, sono:

- 1) Analisi, pianificazione e preparazione.
- 2) Strutturazione ed implementazione delle soluzioni tecnologiche.
- 3) Arruolamento del campione e avvio del progetto pilota.
- 4) Valutazione e validazione di erogazione dei servizi.

Fig. 3  
Fasi del progetto



### 3. Conclusioni

Il progetto si colloca in quello che si propone come uno scenario di Smart Health Care o di Value Based Digital Health Care. In questo contesto gli strumenti digitali assumono sempre maggiore rilevanza non solo sul piano operativo della clinica o della medicina, ma anche su piano di una governance più efficace dei processi sanitari, abilitando una comunicazione diretta con il cittadino che usufruisce di applicazioni e servizi innovativi.

Cloud computing, big data e IoT, nonché le tematiche di cybersecurity e privacy si presentano all'orizzonte di questa sanità "smart". I processi di certificazione dei sistemi per la sicurezza (ISO 27001, cybersecurity act), dei Dispositivi Medici (nuovo Regolamento UE 2017/745), e delle architetture di sistemi e servizi (TOGAF, ITIL, COBIT, ISO 9001) si ritiene possano dare supporto al progetto e alla realizzazione di piattaforme sanitarie "certificate".

## Riferimenti bibliografici

European Commission. eHealth Action Plan 2012-2020: Innovative healthcare for the 21st century | Digital Single Market, 2012. Available from: <https://ec.europa.eu/digital-single-market/news/ehealthaction-plan-2012-2020-innovative-healthcare-21st-century>.

W. Gabriel and R. Åsa. Vision for eHealth 2025 – common starting points for digitisation of social services and health care. eHealth Authority Sweden, 2016. Available from: <https://www.ehalsomyndigheten.se/globalassets/dokument/vision/vision-for-ehealth-2025.pdf>

N. Halsey. International Standards. European Medicines Agency - Science Medicines Health. Available from: [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Presentation/2012/05/WC500126848.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Presentation/2012/05/WC500126848.pdf)

HealthIT.gov. Increase Patient Participation in Care with Health IT | Providers & Professionals, 2014. Available from: <https://www.healthit.gov/providers-professionals/patient-participation>

R. Marsetti, M. Bava et al. – Analisi, progettazione e sviluppo di un'applicazione per Android e IOS per l'IRCCS materno-infantile Burlo Garofolo. Poster presentato a XVI Convegno Nazionale AIIC, Bari 7-9 Aprile 2016

## Autori



**Michele Bava** - [michele.bava@burlo.trieste.it](mailto:michele.bava@burlo.trieste.it)

Laureato in Ingegneria Elettronica, Master in Informatica Medica e Ingegneria Clinica, PhD in Ingegneria dell'Informazione, lavora dal 2003 presso l'Ufficio Sistema Informativo dell'IRCCS "Burlo Garofolo" di Trieste, dal 2009 in qualità di amministratore di sistema e dal 2018 come DPO aziendale. Titolare di diversi progetti di ricerca, svolge attualmente studi nel campo dell'ICT in Sanità, della Telemedicina e della Sicurezza Informatica.

**Valeria Mallozzi** - [valeria.mallozzi@burlo.trieste.it](mailto:valeria.mallozzi@burlo.trieste.it)

Laureanda in Ingegneria Clinica presso l'Università di Trieste, laureata triennale in Ingegneria Biomedica presso l'Università di Pisa, attualmente è ricercatrice presso L'IRCCS "Burlo Garofolo".

**Margherita Dal Cin** - [margherita.dalcin@burlo.trieste.it](mailto:margherita.dalcin@burlo.trieste.it)

Dirigente Medico, Direzione Strategica, IRCCS materno-infantile "Burlo Garofolo".

**Michela Cinello** - [michela.cinello@areasciencepark.it](mailto:michela.cinello@areasciencepark.it)

Ufficio Innovazione e Complessità, Area Science Park, Trieste.

**Sara Zanchiello** - [sara.zanchiello@areasciencepark.it](mailto:sara.zanchiello@areasciencepark.it)

Ufficio Innovazione e Complessità, Area Science Park, Trieste.

## **Social4School: educare alla consapevolezza nei social network**

Livio Bioglio, Sara Capecchi, Valentina Di Noi, Gian Manuel Marino,  
Ruggero G. Pensa, Giulia Venturini

Università degli Studi di Torino, Dipartimento di Informatica

**Abstract.** Social4School è un progetto sviluppato presso il Dipartimento di Informatica (Università di Torino), e cofinanziato dalla Fondazione CRT, allo scopo di aumentare la consapevolezza di giovani e giovanissimi sui meccanismi di interazione e di diffusione delle informazioni nei social network, attraverso un approccio innovativo basato sulla gamification. Il nostro strumento educativo è stato validato nell'ambito di un ampio studio sperimentale che ha coinvolto più di 450 bambini e 22 insegnanti in sette istituti italiani di scuola primaria: i risultati della sperimentazione mostrano che il nostro approccio è stimolante ed efficace nell'aiutare i bambini a scoprire e riconoscere sia le potenzialità che i possibili rischi delle piattaforme social.

**Keywords.** alfabetizzazione digitale, social media, privacy, diffusione della disinformazione, supporto educativo

### **Introduzione**

Diversi studi mostrano che molti utenti di siti web e piattaforme social non sono consapevoli del tema della privacy nella rete e dei rischi in cui possono incorrere nel diffondere informazioni personali, proprie o dei loro amici [Ko-sinski et al. 2013, Furini et al. 2015]: questo problema riguarda sia gli adulti che i minori, ma questi ultimi sono maggiormente colpiti a causa della loro fragilità ed inesperienza.

Un recente sondaggio [IPSOS 2015] condotto dall'istituto IPSOS per Save the Children ha mostrato che, in Italia, gli adolescenti (12-17 anni) sono sempre connessi alla Rete, e conoscono abbastanza bene le regole che regolano la privacy su Internet (51%), ma a loro non importa molto (57%). Inoltre, vivono relazioni virtuali nelle chat room delle applicazioni di messaggistica per smartphone, spesso con persone che non conoscono direttamente (41%): quasi uno su quattro (24%) dichiara di aver inviato messaggi, video o immagini con contenuti sessuali a gruppi con partecipanti sconosciuti, e uno su tre (33%) ha organizzato un incontro con qualcuno conosciuto solo attraverso questi gruppi. Negli ultimi anni, diversi casi di cyberbullismo hanno attirato l'attenzione globale su quanto possa essere rischioso questo tipo di comportamento, in particolare per bambini e adolescenti.

Il problema principale sembra essere la scarsa percezione che i giovani hanno della propria (e altrui) privacy quando si trovano online. Gli adolescenti italiani sembrano ignorare i meccanismi che regolano la diffusione delle informazioni su Internet, specialmente sui social network, e di conseguenza sottovalutano la potenziale diffusione dei loro messaggi, immagini e video nella Rete. Questi numeri sono favoriti dall'uso sempre più diffuso di dispositivi mobili tra i giovani, e dal fatto che imparino ad utilizzarli

principalmente da soli (61%).

Le istituzioni governative stanno provando ad affrontare il problema della comunicazione online proponendo diverse attività, specialmente in ambito scolastico, ma tali attività solitamente si concentrano su tematiche diverse dalla privacy online (come cyberbulismo e hate speech), e sono tendenzialmente condotte in un formato tradizionale, in cui un insegnante svolge in classe una lezione incentrata sull'argomento della comunicazione su Internet, mentre altre forme di apprendimento (come l'apprendimento collaborativo [Dillenbourg 1999], l'active learning [Bonwell et al. 1991] e l'apprendimento collaborativo in rete [Trentin 2010]) si sono dimostrate più efficaci nell'educazione di giovani e giovanissimi.

Social4School [S4S] affronta il problema di migliorare la percezione della privacy e la con-

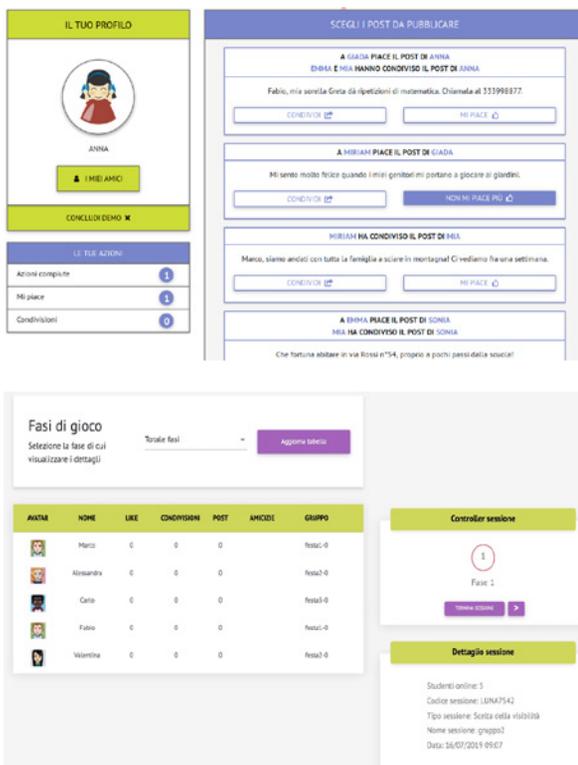


Fig. 1  
Interfaccia sessione di gioco vista dagli studenti (a sinistra) e pannello di controllo a disposizione del docente durante la sessione di gioco (a destra).

sapevolezza degli utenti dei social network mediante un approccio innovativo, basato sulla ludicizzazione. In particolare, proponiamo un'applicazione Web che consente a bambini e adolescenti di sperimentare le dinamiche tipiche della diffusione delle informazioni su un social network attraverso una simulazione interattiva realistica.

Gli studenti che partecipano alla sessione di gioco si connettono all'interno di un piccolo social network locale; attraverso le diverse fasi del gioco, possono scegliere una serie di frasi da pubblicare, e decidere se mettere un like o condividere i messaggi pubblicati dai loro amici (Figura 1, immagine a sinistra). Il nostro gioco ha l'obiettivo di far sperimentare

a giovani e giovanissimi le dinamiche tipiche di un social network in un ambiente simulato e controllato (una sorta di “sandbox”): l'insegnante guida l'intera simulazione attraverso il suo pannello di controllo (Figura 1, immagine a destra), dal quale può anche monitorare l'attività di ciascun partecipante. I giocatori possono così scoprire e toccare con mano la propagazione delle informazioni nella rete di amicizie che si instaura all'interno del social network, ed in questo modo migliorare la loro consapevolezza del fenomeno, la loro percezione dei problemi relativi alla privacy online ed il significato del tema della protezione dei dati, sia propri che dei propri amici.

Al termine della sessione di gioco, l'applicazione calcola una serie di punteggi in base al comportamento tenuto dallo studente, e genera un report interattivo che riporta ogni singola azione effettuata durante la sessione di gioco: queste informazioni possono essere utilizzate dall'insegnante per fornire suggerimenti personalizzati a ciascuno studente, e per stimolare una discussione sul tema della diffusione delle informazioni sui social network al termine della sessione di gioco. L'obiettivo finale della nostra piattaforma è sensibilizzare bambini e giovani sui meccanismi che governano la diffusione di informazioni private nei social network

## 1. Risultati sperimentali

Allo scopo di validare la piattaforma ed il nostro approccio, abbiamo svolto una sperimentazione che ha interessato sette complessi scolastici di scuola primaria in diverse città del Piemonte [Bioglio et al. 2018]. In particolare, la nostra sperimentazione mirava a: i) valutare la consapevolezza dei minori sul tema della privacy nei social media mediante una serie di situazioni specifiche affrontate nella nostra applicazione; ii) valutare la consapevolezza della privacy in situazioni più generali non specificamente affrontate dalla nostra applicazione; iii) valutare le differenze tra il nostro approccio basato sulla ludicizzazione e le altre attività normalmente adottate nelle scuole; iv) ottenere una valutazione generale del nostro approccio e alcuni suggerimenti dagli insegnanti (e dagli studenti) coinvolti nelle nostre attività. L'esperimento ha interessato oltre 450 bambini e 22 insegnanti. I risultati hanno mostrato che il nostro strumento è efficace nel migliorare la consapevolezza dei giovani sulle questioni inerenti alla privacy e alla condivisione di contenuti nei social network.

La sperimentazione è avvenuta come segue. Il sistema al termine del gioco calcola tre punteggi per ogni giocatore: il punteggio attivo (l'impatto che le azioni di un utente hanno avuto sulla privacy degli altri utenti), il punteggio passivo (l'impatto che le azioni di un utente hanno avuto sulla propria privacy) e il punteggio di perdita (che misura in che misura la privacy dell'utente è stata compromessa dalle azioni degli altri utenti). Abbiamo calcolato il valore medio dei tre punteggi per ciascuna classe interessata dalla sperimentazione in due differenti sessioni di gioco: una completamente al buio, senza che gli studenti avessero informazioni sull'attività da svolgere, mentre una seconda qualche tempo successivo alla prima, dopo che gli insegnanti avevano tenuto una lezione sul tema della privacy online, utilizzando la nostra piattaforma per discutere con gli studenti delle loro azioni.

Abbiamo quindi osservato che la distribuzione dei tre punteggi nella seconda sessione di gioco è spostata verso valori più alti, che nel nostro gioco indicano una maggiore attenzione al tema della privacy da parte degli utenti: tale miglioramento è stato osservato in

ogni anno scolastico. Infine abbiamo somministrato dei questionari agli insegnanti sulla loro esperienza con la piattaforma, che hanno confermato la forte motivazione iniziale e la soddisfazione generale per l'attività.

## 2. Sviluppi futuri

Come sviluppi futuri della piattaforma, stiamo estendendo l'approccio Social4School in diverse direzioni: da una parte, intendiamo abilitare la piattaforma al trattamento di altri temi relativi alla comunicazione sul Web (come Cyberbullismo, hate-speech e diffusione di bufale), dall'altra stiamo creando una serie di risorse didattiche di facile accesso per formare il personale educativo all'utilizzo consapevole della piattaforma.

## 3. Ringraziamenti

Il lavoro presentato in questo articolo è stato cofinanziato dalla Fondazione CRT (Fin. N. 2017.2323).

## Riferimenti bibliografici

Kosinski Michal, Stillwell David, Graepel Thore (2013), Private traits and attributes are predictable from digital records of human behavior, PNAS (110:15), pp 5802-5805.

Furini Marco, Tamanini Valentina (2015), Location privacy and public metadata in social media platforms: attitudes, behaviors and opinions, Multimedia Tools Appl. (74:21), pp 9795-9825.

Ipsos Public Affairs, Safer Internet Day Study 2015: i nativi digitali conoscono veramente il loro ambiente? (2015).

Dillenbourg Pierre, Collaborative Learning: Cognitive and Computational Approaches (Advances in Learning and Instruction - 2nd edition (1999), Pergamon Press, Oxford (UK).

Bonwell Charles C., Eison James A., Active learning : creating excitement in the classroom (1991), Clearinghouse on Higher Education, George Washington University, Washington, DC (USA).

Trentin Guglielmo, Networked Collaborative Learning (2010), Chandos Publishing, Cambridge (UK).

S4S, piattaforma web disponibile all'indirizzo web [www.social4school.net](http://www.social4school.net).

Boglio L., Capecchi S., Peiretti F., Sayed D., Torasso A., Pensa R.G. A Social Network Simulation Game to Raise Awareness of Privacy among School Children. IEEE Transactions on Learning Technologies. 2018, 14 pagine. IEEE. Disponibile online

## Autori



Livio Bioglio - [livio.bioglio@unito.it](mailto:livio.bioglio@unito.it)

Livio Bioglio si è laureato in Informatica nel 2009 presso l'Università di Torino, e ha ricevuto il dottorato di ricerca in Informatica nel 2013 presso la stessa istituzione. Nel 2013 è stato Post-doc presso INSERM, Parigi (Francia), dove ha lavorato su sistemi complessi, epide-

miologia, e modelli computazionali per l'analisi della diffusione di malattie. Dal 2016 è Post-doc presso l'Università di Torino, dove studia la diffusione delle informazioni nei social network.

**Sara Capecchi** - [sara.capecchi@unito.it](mailto:sara.capecchi@unito.it)

Sara Capecchi ha ricevuto nel 2006 il dottorato in Informatica presso l'Università di Firenze. Dal 2005 al 2010 è stata post-doc presso le Università di Catania e di Torino, e dal 2010 è Ricercatrice presso il Dipartimento di Informatica dell'Università di Torino. I suoi principali interessi di ricerca includono modelli concettuali per sistemi di fiducia e reputazione e analisi statiche di sistemi distribuiti con particolare attenzione alla perdita di informazioni e al controllo degli accessi.



**Ruggero G. Pensa** - [ruggero.pensa@unito.it](mailto:ruggero.pensa@unito.it)

Ruggero G. Pensa si è laureato nel 2003 in Ingegneria Informatica presso il Politecnico di Torino e ha ricevuto il dottorato di ricerca in Informatica presso l'INSA di Lione nel 2006. Dal 2011 è Ricercatore presso il Dipartimento di Informatica dell'Università di Torino. I suoi principali interessi di ricerca includono data mining, machine learning, algoritmi per la tutela della privacy per la gestione e analisi dei dati, analisi dei social network e analisi di dati spazio-temporali.

**Valentina Di Noi** - [valentina.dinoi@unito.it](mailto:valentina.dinoi@unito.it)

Valentina Di Noi si è laureata nel 2017 al D.A.M.S in nuovi media presso l'Università di Torino. Nello stesso anno ha ottenuto un diploma di specializzazione in grafica 3D. Nel 2018 frequenta un Master in Progettazione e Management del multimedia per la comunicazione. Dal 2017 a oggi ha lavorato principalmente nel settore della grafica 3D, del Web Design e dell'E-learning. Dal 2017 è borsista presso il Dipartimento di Informatica di Torino.



**Gian Manuel Marino** - [gian.marino@edu.unito.it](mailto:gian.marino@edu.unito.it)

Gian Manuel Marino ha conseguito la Laurea Triennale in Informatica nel 2017 presso l'Università degli Studi di Torino. Attualmente sta concludendo gli studi per la Laurea Magistrale presso la stessa facoltà. Dal 2017 al 2019 è stato borsista presso il dipartimento di Informatica dell'Università degli Studi di Torino.

**Giulia Venturini** - [giulia.venturini@unito.it](mailto:giulia.venturini@unito.it)

Giulia Venturini si è laureata in Scienze della Comunicazione presso l'Università degli Studi di Torino, con una tesi sull'accessibilità web. Attualmente ha una borsa di ricerca presso l'Università di Torino, in cui si occupa di progettazione e produzione di contenuti multimediali per la formazione a distanza in ambito tecnico-scientifico.



# Nuove forme di fruizione del digitale. Il repository Byterfly: la conservazione e il riuso con l'open source, access e data

Giancarlo Birello, Anna Perin

CNR-IRCrES, Ist. di Ricerca sulla Crescita Economica Sostenibile, Moncalieri (TO)

**Abstract.** Questo paper descrive l'esperienza maturata dall'Ufficio IT e dalla Biblioteca CNR-IRCrES nello sviluppo di una biblioteca digitale, Byterfly, con requisiti di open source per architettura e software, open access per contenuti e open data per filosofia di conservazione, fruizione e riuso.

**Keywords.** Biblioteca digitale, repository, open source, open access, open data

## Introduzione

Un repository è una biblioteca digitale con una raccolta mirata di oggetti che può includere testi, immagini, file audio e video, tutti archiviati in collezioni e associati a metadati che forniscono informazioni sulla risorsa. Tali oggetti devono essere ricercabili, accessibili e fruibili. L'ufficio IT e la biblioteca del CNR-IRCrES hanno sviluppato e gestiscono alcuni repository dal 2012, Byterfly è uno di questi (Birello et al., 2018) con requisiti di open source per architettura e software, open access per contenuti e open data per filosofia di conservazione, fruizione e riuso.

## 1. La biblioteca digitale

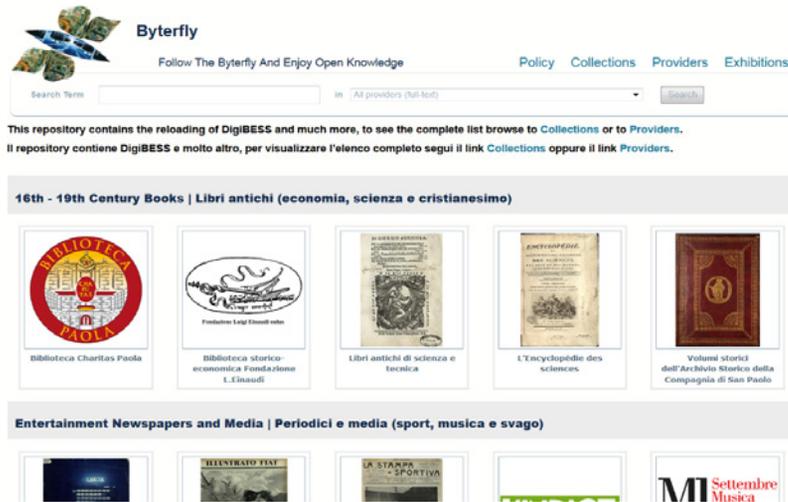
Per la sua realizzazione è stato utilizzato Fedora Repository, uno dei software più adottati a livello mondiale per la gestione delle risorse digitali, robusto, persistente e in grado di gestire big data. Per il front-end è stato utilizzato Dupal CMS e Islandora per l'accesso e la presentazione degli oggetti digitali, dotato di moduli espandibili e personalizzabili per tutti i tipi di media. Per le funzioni di ricerca è stato utilizzato Solr. I metadati adottati sono i Dublin Core. L'architettura è OAI-PMH e IIIF Image and Presentation API 2.1 compliant, il che consente la visibilità e la disseminazione sia attraverso meta-repository (Worldcat per esempio), ma anche semplicemente dai motori di ricerca tradizionali. Attualmente Byterfly ospita ventuno fornitori di contenuti (provider). Ogni provider può avere una o più collezioni. Le collezioni presenti (65) sono molto diversificate, ma consideriamo questo un punto di forza e non un limite.

Nella biblioteca digitale è depositato molto materiale sulle scienze sociali ed economiche del Piemonte inclusi volumi dell'Università di Torino (appartenenti alla biblioteca Bobbio e alla biblioteca di Economia e Management), della Fondazione Einaudi,

periodici del Gruppo Fiat (FGA Automotive), della Camera di Commercio di Torino, del Museo dell'Automobile di Torino, libri d'arte (provider Fondazione 1563 per Arte e Cultura di Torino). Inoltre, ospitiamo libri antichi (sec. XVI-XVIII) di teologia e religione (provider: Ordine dei Minimi di San Francesco di Paola), volumi antichi di scienza e tecnica (provider: CNR-IRCrES) e materiali, libri e immagini del festival musicale internazionale MITO Settembre Musica organizzato ogni anno a Torino e Milano. I documenti sono in italiano, latino (libri antichi), inglese, francese, spagnolo e tedesco.

Per meglio orientare l'utente alcune collezioni sono state visivamente raggruppate nella homepage in 5 macro aree: Libri antichi, periodici e media, riviste di economia e impresa, riviste di scienze sociali, libri moderni di economia e scienze sociali. Attualmente sono depositate ed accessibili più di 1.300.000 pagine, in continua espansione.

Fig. 1  
Home page Byterfly  
www.byterfly.eu



## 2. L'architettura

L'architettura si basa su una piattaforma virtualizzata che utilizza KVM come hypervisor in combinazione con OpenVSwitch per il networking e iSCSI con multipath per gestire la SAN (Storage Area Network).

L'architettura attuale richiede tre macchine virtuali:

- il back-end, che include il repository per gestire gli oggetti digitali;
- il server IIIF (International Image Interoperability Framework) per il trattamento delle immagini;
- il front-end, che ospita l'interfaccia per l'accesso pubblico agli oggetti e ai dati del repository.

Uno storage altamente affidabile (SAN) completa l'architettura fornendo al repository un ampio spazio per la conservazione degli oggetti e all'hypervisor lo storage per i backup delle macchine virtuali.



Fig. 2.  
Architettura Byterfly  
a tre macchine  
virtuali

Stessa architettura è utilizzata anche per il repository ASA – Archivio Studi Adriatici (Armeli Minicante et al., 2017).

La macchina virtuale di back-end ospita Fedora Repository, BlazeGraph e Solr, ognuno dei quali viene eseguito in un contenitore servlet Java separato: i primi due, in un paio di istanze di Tomcat, mentre Solr è in esecuzione in un contenitore servlet Jetty incluso nella distribuzione di Solr. Fedora Repository gestisce l'attività di conservazione dei contenuti della Byterfly e include le API per l'ingesting programmabile, la descrizione semantica delle relazioni tra gli oggetti e la loro gestione basata su modelli. BlazeGraph è il gestore del triple-store delle relazioni tra gli oggetti del repository. Infine, Solr è il motore di ricerca e indicizzazione di tutti gli oggetti del repository. La configurazione è personalizzata per indicizzare anche i full-text dei volumi oltre a tutti gli elementi Dublin Core dei metadati al fine di fornire i campi da utilizzare nel front-end Islandora, nelle pagine di ricerca e visualizzazione.

Il server IIIF Cantaloupe fornisce l'accesso alle immagini conservate nel repository tramite le Image API 2.1 dello standard. Può accedere sia alla versione JPEG2000 che TIFF delle immagini per poter fornire al front-end i riquadri opportunamente scalati richiesti dai visualizzatori OpenSeaDragon e Internet Archive Bookreader. Cantaloupe è un generatore di derivative ad alte prestazioni con potenti funzionalità di cache ed estremamente versatile, in grado di fornire immagini a più coppie repository/front-end contemporaneamente.

Server front-end "Open-access" significa poter accedere a tutti gli oggetti digitali e ai dati conservati nel repository in modo semplice. Questo è uno dei motivi principali per cui abbiamo scelto Islandora molti anni fa e abbiamo confermato la scelta anche per Byterfly. Il framework Islandora è integrato in Drupal CMS, supportato da una community dinamica ed estesa. Reso inoltre interessante per il supporto IIIF e l'uso intensivo di Views personalizzabili per visualizzare nel front-end collezioni, oggetti e metadati in base alle più svariate esigenze. Non ultima la recente implementazione delle IIIF Presentation API 2.1 tramite un modulo di Islandora che rendono i contenuti della Byterfly compatibili con gli standard richiesti da Europeana.

### **3. Le policy**

Le policy adottate per il repository sono disponibili in una pagina web sul portale e si basano sulle indicazioni fornite da OpenDOAR (The Directory of Open Access Repositories) che, con un semplice strumento online, aiuta a formulare e presentare le policy del proprio repository con opzioni raccomandate minime per la conformità di adesione al movimento Open Access. Le policy coprono tutti gli aspetti del repository, in particolare: Policy dei metadati, dei dati, dei contenuti, di sottomissione e di conservazione.

### **4. Mostre digitali**

È stato inoltre realizzato un sito di mostre digitali (exhibitions) che consente una fruizione trasversale del materiale contenuto nella biblioteca digitale. Attualmente è presente la mostra “1919. Cronache dalla Storia: immagini e documenti d’archivio. Dal quotidiano alla voglia di cambiamento”. Realizzata in occasione del centenario della fine della Grande Guerra (1915-1918), offre un percorso organizzato attorno a quattro grandi temi: “La ripresa delle attività”, “Economia e politica”, “La donna e la società”, “Lo sport e lo svago”.

### **5. Sviluppi futuri**

Da diversi anni collaboriamo allo sviluppo del software per la realizzazione di repository con alcune comunità dedite all’open source, negli ultimi mesi è nata una stretta collaborazione con il Metropolitan New York Library Council (metro.org) che porterà nei prossimi mesi al rilascio di una nuova architettura, semplificata e flessibile che prevede metadati dinamici, altamente integrata con Drupal e che sfrutta a pieno le potenzialità del IIIF.

### **6. Conclusioni**

Obiettivo della Byterfly è mettere al centro ogni collezione ospitata e, anche se piccola, sfruttare i vantaggi dell’architettura generale. Massima attenzione alla fruibilità degli oggetti, semplicità di utilizzo della piattaforma da parte dell’utente e preservazione a lungo termine dei materiali contenuti.

### **Riferimenti bibliografici**

Birello G., Perin A. (2018) Follow the Byterfly and enjoy open knowledge, Rapporto Tecnico CNR-IRCrES, n. 6/2018. <http://dx.doi.org/10.23760/2421-5562.2018.006>

Armeli Minicante S., Birello G., Sigovini M., Minuzzo T., Perin A., Ceregato A. (2017): Building a Natural and Cultural Heritage Repository for the Storage and Dissemination of Knowledge: The Algarium Veneticum and the Archivio di Studi Adriatici Case Study, *Journal of Library Metadata*, Volume 17, Issue 2, Pages: 111-125, DOI:10.1080/19386389.2017.1355165

## Autori



**Giancarlo Birello** - [giancarlo.birello@ircres.cnr.it](mailto:giancarlo.birello@ircres.cnr.it)

È System e Network Manager presso l'ufficio IT di CNR-IRCrES, referente regionale per l'infrastruttura di rete CNR Piemonte e Access Point Manager per il GARR. Ha iniziato ad usare programmi open source per la conservazione e presentazione di oggetti digitali dieci anni fa, ha recentemente implementato la biblioteca digitale Byterfly. È coinvolto in alcune comunità internazionali per lo sviluppo dei software utilizzati.

**Anna Perin** - [anna.perin@ircres.cnr.it](mailto:anna.perin@ircres.cnr.it)

È responsabile della Biblioteca CNR-IRCrES dove svolge tutte le attività di back office e front office per i propri utenti, servizi di reference e di orientamento per l'utilizzo delle risorse della biblioteca e document delivery. Si occupa della realizzazione e gestione dei siti web di Istituto. Ha sviluppato competenze sui repository di ultima generazione occupandosi in particolare di metadati, policy, usabilità e fruibilità per l'utente.



# Adalitix: a Web Platform for Agricultural Data Analytics

Luca Coviello, Marco Cristoforetti, Cesare Furlanello

Fondazione Bruno Kessler

**Abstract.** Digital technologies ignited a revolution in the agrifood domain known as smart farming. In this work we present Adalitix, a cloud-based web platform for agricultural data analytics (including: collection, visualization and predictive modelling) for better crop management. We also describe two real smart farming cases, where Adalitix is used by vineyard technicians to estimate grape quality parameters and predict yield estimation directly in field. Data were acquired with a 2 Mpx smartphone camera and a handheld portable spectrometer with non-invasive techniques.

**Keywords.** Smart Farming, Agricultural Data, Web Platform, Deep learning, Complex Analytics

## Introduction

The continuous development and adoption of connected IoT devices is impacting all aspects of our day life and environment. The aim is to build a smart society that takes decision based on the insights generated by analysing data collected with low cost sensors in different domains. With the always growing demand of food (+50% between 2012 and 2050), the few opportunities left to expand agricultural areas and the significant decline in yield growth (FAO, 2017), it is crucial to develop new technologies and infrastructures that help to produce more and safer food with less resources. Smart farming represents a unique opportunity for helping farmers to measure, monitor and analyse precisely the status of their crops in real time. Significant factors that slow down the adoption of digital technologies in the agrifood industry are the reluctance of growers to employ IT systems in field given their cost, resource demand and difficulty to use. In this work we propose Adalitix (Fig. 1), a web based solution to store, visualize and analyse agricultural data collected with in field sensors. Furthermore, it allows farmers to leverage cloud computing resources to automatically apply Artificial Intelligence (AI) models, e.g. deep learning, to forecast future crop development and improve decision making processes. Notably, Adalitix makes easy to publish the generated datasets, e.g. for research or data monetization purposes, with standardized metadata. Two use cases developed for viticulture will illustrate the benefits of Adalitix for agricultural technicians, enabling the adoption of AI techniques in a consistent and efficient framework.

## 1. Adalitix

### 1.1 Field Data Measurements

Agricultural field measurements can be recorded with different instruments or sensors and in different formats (e.g. images, temperatures, humidity, etc.). Adalitix was desig-

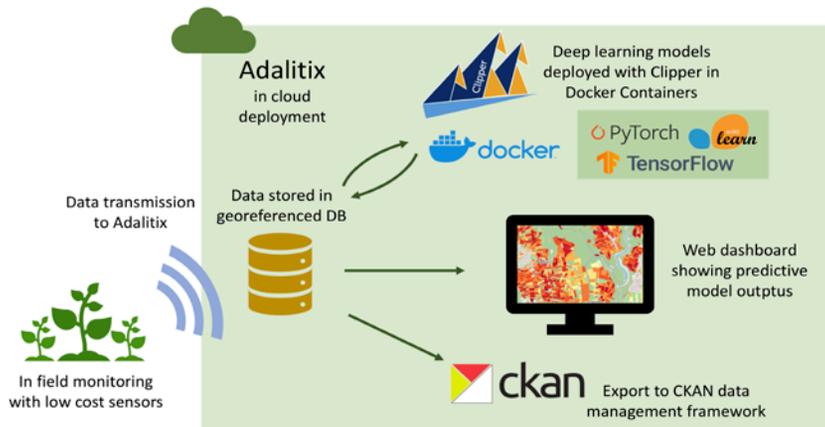


Fig. 1.  
Adalitix  
architecture  
components  
overview

ned to store various types of data (e.g. images, time series, geoTIFFs) with a particular focus on the spatial and temporal resolution, which directly impacts the analysis results. In Adalitix, data are organized into datasets and can be visualized or given as input to the analysis pipelines deployed on the system. Results are then stored and made available on the web platform.

## 1.2 Web Dashboard

The stored information can be visualized in the web dashboard to monitor field conditions and crop status and development. Moreover, data can be represented over the spatial and temporal axes to analyze short or long term changes. To this aim Adalitix displays the collected information on a map with the option to focus on particular areas or time spans. A dashboard enables selection and visualization of outputs generated from the analysis pipelines available on the platform. In particular, Adalitix functions are being designed to forecast the phenological development of plants and predict crop yield given other environment variables.

## 1.3 AI Analytics Pipelines

AI is a key enabling technology for extracting valuable insights from data, although requiring massive data and significant computing power for training models. The computational resources required at inference (use of trained models) are instead usually limited: models can be run on field on edge computing devices, smartphones or cloud systems at the cost of data transmission. Deep learning models and complex analytics are deployed in Adalitix with Clipper, an online prediction serving system (Crankshaw, Wang 2017). Clipper deploys every model and pipeline as an independent docker container with the required dependencies, accessible through a RESTful API. On the Adalitix web platform, users can upload pipelines on the stored datasets, manage results and make them available through the dashboard.

## 1.4 CKAN metadata and data indexing system

To simplify the data access and search process we adopted CKAN (Comprehensive

Knowledge Archive Network), an open source data management system that allows to publish, share, search and use data. CKAN provides a metadata repository and indexing system to easily retrieve, publish and export datasets for download or integration on other CKAN instances. Moreover, CKAN exposes its content and functionalities through RESTful APIs following the specifications defined by the Agency for Digital Italy (AgID) as national standard for data sharing between administrations. The CKAN Adalitix instance allows users to export its catalogue as a collection of metadata and descriptors.

## 2. Use Cases

Adalitix has been employed for noninvasive grape yield prediction and grape quality estimation, based on deep learning algorithms. Both classes of models are made available to vineyard technicians through the combined use of a smartphone and a handheld portable spectrometer directly in the field. Other IoT options are also available, e.g. connectivity by LoRaWan. For yield estimation, we adapted CSRNet (Li et al. 2018), a deep learning algorithms developed for the crowd counting context to estimate the number of grapes present in an image. The model reached an overall mean average percentage error MAPE = 10% (~17000 grape berries) with images taken from a medium distance. For quality estimation we developed a deep learning model applied to spectral signals (Zhang et al. 2019) (Ni et al. 2019) for sugar content estimation in grapes from ~2000 samples, reaching an  $R^2 = 0.83$  and a mean average error MAE = 0.77 brix degrees.

## 3. Conclusions and Future Work

Agriculture provides an elective application domain for AI combined with IoT technologies, mostly requiring low cost sensors, making the development of data platforms a crucial to increase productivity, value and quality of production. In this work we presented Adalitix, a web platform for storing, visualizing and analysing data collected with in field instruments connected to the Internet. The two AI solutions enabled by our platform for wine production illustrate how predictive models can introduce in smart farming ecosystems new tools to estimate quantity and quality. Further developments include upscale of data collection with low cost network technologies (e.g. Long Range Low power networks), enabling the use of Adalitix solutions with georeferenced time series data from smart devices deployed in remote areas with low Internet coverage.

## 4. Acknowledgements

The work presented here is sponsored by the Consortium GARR with the “Orio Carlini” research scholarship. FBK acknowledge support by CAVIT. Initial work has been developed within the FBK WebValley Data Science Project 2016 and 2017. Collaboration with CODIPRA Trento and MPA Solutions is also acknowledged.

## References

- FAO. (2017). The future of food and agriculture. Trends and challenges.
- Crankshaw, D., Wang, X., Zhou, G., Franklin, M. J., Gonzalez, J. E., & Stoica, I. (2017).

Clipper: A low-latency online prediction serving system. In 14th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 17), pp. 613-627.

Li, Y., Zhang, X., & Chen, D. (2018). Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1091-1100.

Zhang, X., Lin, T., Xu, J., Luo, X., & Ying, Y. (2019). DeepSpectra: An end-to-end deep learning approach for quantitative spectral analysis. *Analytica Chimica Acta*.

Ni, C., Wang, D., & Tao, Y. (2019). Variable weighted convolutional neural network for the nitrogen content quantization of Masson pine seedling leaves with near-infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* (209) pp. 32-39.

## Authors



**Luca Coviello** - [coviello@fbk.eu](mailto:coviello@fbk.eu)

Luca Coviello received the M.Sc. in Data Science in 2018 from the EIT Digital Master School. He has been working at FBK since 2014, focusing on the application of machine learning to bioimaging and the development of cloud-based frameworks for large scale dataset analysis. He is now a GARR fellowship student at FBK, working on applying artificial intelligence to agriculture as a low-cost solution for improving crop yield.

**Marco Cristoforetti** - [mcristofo@fbk.eu](mailto:mcristofo@fbk.eu)

Marco Cristoforetti graduated in theoretical physics in 2004 at the University of Milan and received his PhD in Physics at the University of Trento in 2007. Until 2014 he devoted himself to the study of non-perturbative QCD. Data scientist at FBK since 2014 applying Machine Learning and Deep Learning to data from heterogeneous sources such as social networks, agriculture and high energy physics. Member of the CERN ATLAS experiment.



**Cesare Furlanello** - [furlan@fbk.eu](mailto:furlan@fbk.eu)

Cesare Furlanello is a data scientist, head of the FBK Unit for Predictive Models in Biomedicine and Environment. He is an expert in machine learning and bioinformatics applied to multi-modal health data (omics, bioimaging, medical phenotypes) with Deep Learning. Full professor habilitation in BioEngineering. Bioinformatics collaborator of US FDA/NCTR, RIKEN, and adjoint Faculty of the Wistar Institute in Philadelphia, he is the Past President of the MAQC Society for reproducibility and quality control over massive data.

# How to leverage Measurement Lab Internet metrics to get useful insights through Data Science

Stefania Delprete<sup>1</sup>, Gianni Spalluto<sup>2</sup>, Christian Racca<sup>1</sup>

<sup>1</sup>TOP-IX Consortium, <sup>2</sup>Eutelsat Communications

**Abstract.** Measurement Lab is an open source project that provides data on Internet performance measurements. TOP-IX Consortium decided to leverage the open data provided by M-Lab to explore the impact of its network on the territory and the factors that might influence the performances of the connectivity providers among its members, in particular FWA operators. This paper presents the data acquisition process and three case studies related to commercial plans comparison, coverage and weather, including the advantages and obstacles of the data used and approach presented.

**Keywords.** Internet, data science, open data, metrics, performances, weather.

## Introduction

How can we leverage the biggest open dataset on Internet metrics to create reports for specific providers? This paper presents a viable approach to acquire, manage and explore the data, and comparing known information or aggregating other external sources.

If validated, this approach might be used to enrich the information that connectivity providers extract from their own networks to offer an objective multi-operator analysis and to represent a good tool to verify (thanks to the dataset openness) the reports created by the operators themselves or other authorities.

## 1. TOP-IX Consortium and its interest in Internet metrics

TOP-IX (TOriNO Piemonte Internet eXchange) is a non-profit consortium founded in 2002 with the aim of creating and managing a neutral hub for the exchange of Internet traffic in North-West Italy. Furthermore, TOP-IX promotes and supports, through the Development Program (DP), technological, engagement and training projects based on the Broadband Internet, Data and People. These activities synergistically promote the growth of the territory.

According to the typical role of an Internet Exchange, TOP-IX operates at layer 2 in OSI model and it is not allowed, by law, to analyze the content of the exchanged traffic over the platform. Furthermore TOP-IX, in its role as a neutral hub, typically does not have the chance to study and track the impact of network events (port saturation, technical faults, other phenomenon) on final end-users.

This study is aimed at expanding the perspective by TOP-IX on its network backbone and at investigating the feasibility of using third-party datasets to provide value added

“services” (such as report, dashboard or more advanced tools) to the stakeholders active on the IX platform.

## 2. Measurement Lab project and data availability

TOP-IX, since the beginning, demonstrated a strong interest in data collected by network performance analysis tools. During the years TOP-IX activated valuable collaborations with Ookla (1) and Measurement Lab (2) in order to host their tools for performance monitoring.

We started analysing data from the Speedtest by Ookla and, even if it was an insightful starting point to analyse Internet metrics, it couldn't be compared with the amount of variables offered by M-Lab.

Measurement Lab, founded in 2009 to offer a better solution to Internet measurement experiments, collects when available more than 150 variables (3) including: log time, geolocation, browser and operating system, and Internet metrics such as Round-trip delay time (RTT). This kind of approach and transparency attracts interest in analysing data often not available elsewhere.

M-Lab provides a detailed documentation to express the most common network metrics (latency, download and upload speed) applying known formulas in literature on a set of parameters.

For example, from the raw data of a speed test, the download speed in Mbps is expressed by the formula (4):

$$8 * (\text{web100\_log\_entry.snap.HCThruOctetsAcked} / (\text{web100\_log\_entry.snap.SndLimTimeRwin} + \text{web100\_log\_entry.snap.SndLimTimeCwnd} + \text{web100\_log\_entry.snap.SndLimTimeSnd}))$$

We would like to thank M-Lab team, in particular Chris Ritzo and Roberto D'Auria, for their assistance in the interpretation of the data acquisition processes and parameters.

## 3. Tools for gathering data and managing analysis

Measurement Lab uses Google BigQuery to store the tests results in different Tables. The download Table and upload Table contain test results already filtered based on outliers and possible mistakes. Through a whitelisted email address we were able to access Google BigQuery using standard SQL and explore viable solutions to gather the dataset needed for our research.

After exploring the Python BigQuery library, we did a preliminary analysis. We used a Jupyter Notebook to import the library and directly generate the Pandas DataFrames from a SQL query in order to explore and collect the results in one place.

For our purposes we also decided to write a script in Python, with the assistance of Massimo Santoli, to uniquely match the IP addresses in M-Lab tests and TOP-IX members' ASNs (Anonymous System Numbers).

Through this process we were able to have a clean dataset in .csv format related to selected members of TOP-IX Consortium.

## 4. Case studies

This section presents the on-going case studies developed by TOP-IX using the dataset and the approach described above. Currently, we decided to focus our attention on the FWA (Fixed Wireless Access) members of the Consortium. The analyses were performed in Jupyter Notebooks using Python and its computational libraries (NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn).

### 4.1 Profiling connectivity providers and studying correlation between network performance metrics and commercial public plans

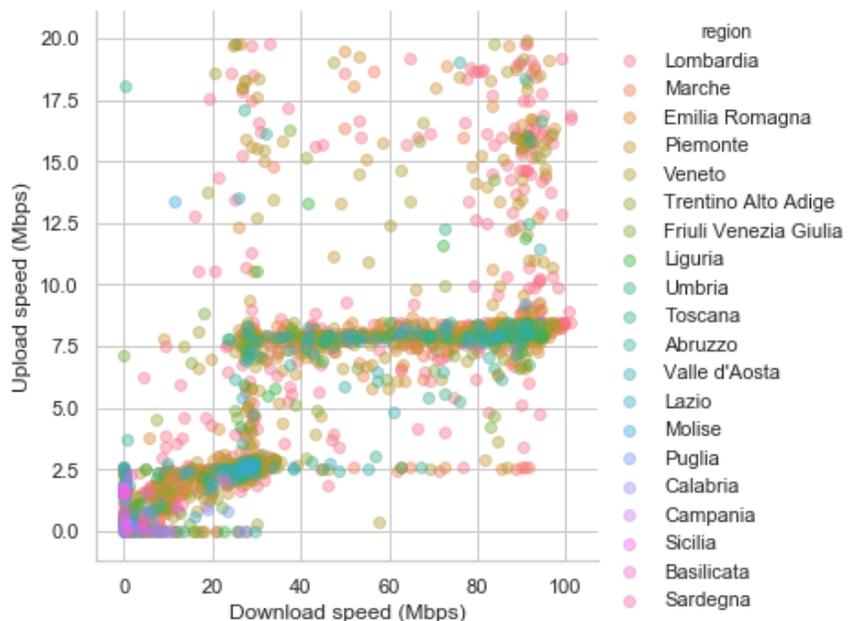
The Data Science project started by exploring different FWA operators from an aggregated point view, to understand which parameters could be addressed and to verify the amount of data and test performances over the years.

A small number of providers was selected for a deeper exploration. This analysis has been useful to compare commercial public plans (openly published by operators) and performances measured through the speed tests. Fig. 1 is an example of how much the speed test can mirror the offerings by noticing the visible horizontal lines on precise Upload speed values. This correlation has been studied by provider and regional area, and aggregating by city and IP.

### 4.2 Role of TOP-IX network as “digital enabler”

TOP-IX has a limited visibility on the network and rarely it can observe the actual impact on the final users. Data from this experiment represents a good proxy for a wider study aimed at exploring, for example, the overall coverage (in terms of connectivity), by the

Fig. 1  
Correlation of upload and download speed for a provider over different regions and focusing on download speed values below 110 Mbps and upload speed values below 20 Mbps



network operators interconnected to TOP-IX.

The analysis started with an exploration to avoid outliers and understand the most covered areas macroscopically. In a second phase we used QGIS (see an example in Fig. 2) to visualise the data using different layers and parameters to make more visible the area with the best performances.

Interns involved in this study: Domenico Gallo, Christian Bellafemmina

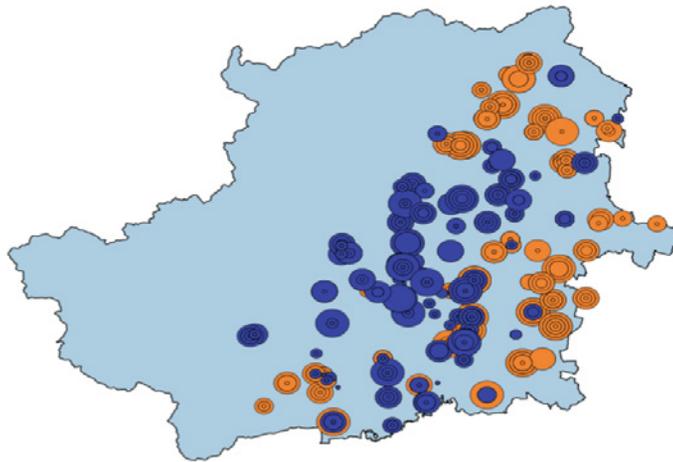


Fig. 2  
Two different providers  
mapped in Turin area,  
the radius of each point  
is proportional to the  
download speed value

### 4.3 Correlation between performance, weather conditions and other phenomena

The website of the Regional Environmental Protection Agency, ARPA (5), offers, under request, the possibility to access the latest weather data (temperature, precipitation, humidity, wind speed, intensity and direction) organized in .csv datasets. Fig. 3 shows part of the experimental approach comparing one of the weather variable (wind intensity) over time.

The measurements of the selected meteorological variables are organized by daily time bands. The aggregation with available Measurement Lab data allows us to test the possible correlation between network performances and weather conditions. Unfortunately the values of the Pearson correlation coefficient were not relevant to make concrete assumptions for the cases studied until this point.

Interns involved in this study: Gianni Spalluto, Paul Cristian Prisecaru, Adriana Muscau.

## Conclusions and further development

The methodology and the use-cases presented above open new possibilities to mix data science and networking. The goal is to get insights from the raw data collected by Measurement Lab as a possible integration and an alternative to “traditional” network monitoring system and official reports not released in real time.

Aggregating datasets from multiple sources might give a better awareness of users’ behaviour to improve network reliability and to offer more customized services from the vendor point of view.

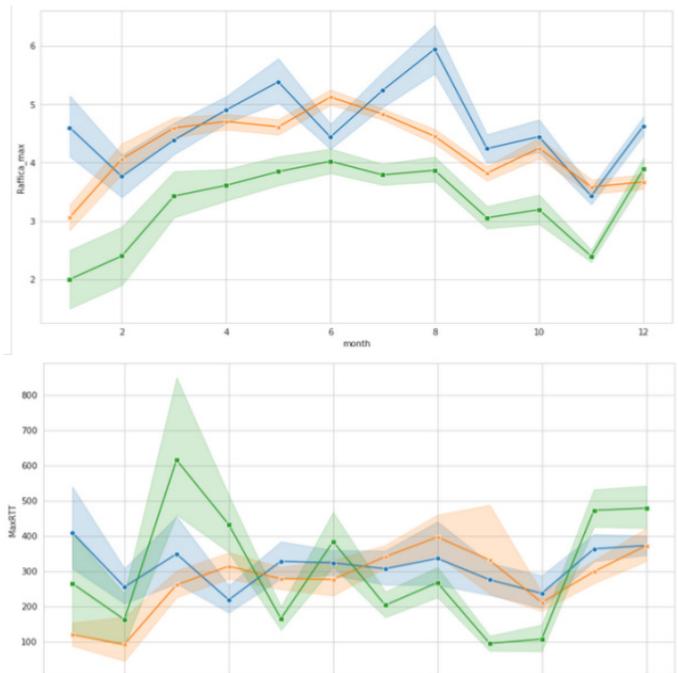


Fig. 3  
Example of trends of  
wind intensity and  
latency for three  
different cities over  
twelve months.

On the other hand we are still working to deal with the imprecision and the gaps in the datasets. For instance both M-Lab and Speedtest by Ookla are using MaxMind as a geolocation reference, this kind of process can lead to incorrect conclusions regarding behaviours based on location.

Furthermore, regarding the integration with weather condition data, for some urban and rural areas, the information is missing or incomplete, or accessible only through private services.

Future developments will include a detailed and precise exploration and pre-processing of the dataset to avoid any misleading conclusions, and further analysis of less common metrics incorporating Machine Learning algorithms to test the possibility to predict performances and stability of the system.

## References

Speedtest by Ookla <https://www.speedtest.net>

Measurement Lab <https://www.measurementlab.net>

“BigQuery Schema” by Measurement Lab

<https://www.measurementlab.net/data/docs/bq/schema>

“Calculating Common Metrics for NDT Data” by Measurement Lab

<https://www.measurementlab.net/data/docs/bq/ndtmetrics>

Weather dataset by ARPA Piemonte <http://www.arpa.piemonte.it>

## Authors



**Stefania Delprete** - [stefania.delprete@top-ix.org](mailto:stefania.delprete@top-ix.org)

Data Scientist, Python tutor, and BIG DIVE co-organizer at TOP-IX with a background in Theoretical Physics and strong interests in Neuroscience, Human Rights, and Social Change. Stefania has been involved in projects in Italy, Germany and UK on productivity and self-improvement, social aid, and open source as a speaker and tech tutor. She volunteers for PyCon/Pydata conferences, and runs local chapters and events for Mozilla and Rust, Effective Altruism, and MathsJam.

**Gianni Spalluto** - [gianni.spalluto1991@gmail.com](mailto:gianni.spalluto1991@gmail.com)

Graduated in Sociology and Applied Social Science, passionate in data analysis and convinced supporter of the Big Data revolution as a catalyst for Social Change and a more equitable society.

Gianni attended IFTS Big Data course and, thanks to internship in TOP-IX, started using Python tools in Data Science projects. He's currently working at Eutelsat group as a human resources analyst where he uses the large amount of data available in the department to support company welfare policies and work life balance of employees.



**Christian Racca** - [christian.racca@top-ix.org](mailto:christian.racca@top-ix.org)

After graduating in Telecommunication Engineering at Politecnico di Torino, Christian joined TOP-IX, working on data streaming and cloud computing, and later on web startups. He has mentored several projects on business models, product development and infrastructure architecture and cultivated relationships with investors, incubators, accelerators and the Innovation ecosystem in Italy and Europe. Currently Christian manages the TOP-IX BIG DIVE program aimed at providing training courses for data scientists, companies, organizations and consultancy projects.



## **TDM: un sistema aperto per l'acquisizione di dati, l'analisi e la simulazione su scala metropolitana**

Fabio Bettio<sup>1</sup>, Giovanni Busonera<sup>1</sup>, Marco Cogoni<sup>1</sup>, Roberto Deidda<sup>2</sup>, Mauro Del Rio<sup>1</sup>, Massimo Gaggero<sup>1</sup>, Enrico Gobetti<sup>1</sup>, Simone Leo<sup>1</sup>, Simone Manca<sup>1</sup>, Marino Marrocu<sup>1</sup>, Luca Massidda<sup>1</sup>, Fabio Marton<sup>1</sup>, Marco Enrico Piras<sup>1</sup>, Luca Pireddu<sup>1</sup>, Gabriella Pusceddu<sup>1</sup>, Alessandro Seoni<sup>2</sup>, Gianluigi Zanetti<sup>1</sup>

<sup>1</sup>Centro di Ricerca, Sviluppo e Studi Superiori in Sardegna

<sup>2</sup>Università di Cagliari

**Abstract.** Il progetto TDM studia e sviluppa nuove tecnologie abilitanti e soluzioni verticali in ambito Smart Cities, sperimentandole nell'area metropolitana della Città di Cagliari. Il design scalabile, l'utilizzo di standard per le Smart Cities (OASC/FIWARE) e la realizzazione di un dispositivo per l'integrazione di sensori diversi (l'Edge Gateway) fanno sì che le soluzioni realizzate nel contesto del progetto TDM possano essere estese ad altre realtà urbane. In questo articolo illustriamo l'architettura generale del sistema, il sottosistema dedicato alla sensoristica e le prime applicazioni in campo energetico e meteo-ambientale.

**Keywords.** smart-cities, fiware, edge-computing, energy-awareness, nowcasting

### **1. Architettura generale**

Il progetto TDM combina lo studio e sviluppo di tecnologie abilitanti, l'applicazione a tematiche strategiche e la sperimentazione a scala metropolitana nella Città di Cagliari, al fine di definire soluzioni scalabili e best practice generali da utilizzare in altre realtà urbane ampie o densamente popolate. A tal fine, TDM integra sorgenti di informazioni digitali eterogenee e metodi di analisi e simulazione per la realizzazione di modelli digitali multi-scala di aree urbane metropolitane. Dati corrispondenti a livelli diversi di informazione sono quindi organizzati, dal punto di vista logico, in strutture opportune come serie temporali per misure da sensori "puntuali", e strutture volumetriche per campi continui su regioni dello spazio. I formati di rappresentazione dei dati adottate sono standard e best practice del dominio Smart Cities, nello specifico i Modelli di Dato Armonizzati di FIWARE, usati in ambito OASC per i dati da sensore, e lo standard internazionale CF (Climate and Forecast) per i dati provenienti da simulazioni e acquisizioni radar.

L'architettura di TDM è strutturata su tre macro-blocchi dedicati all'acquisizione distribuita dei dati, alla loro elaborazione e, infine, alla presentazione e distribuzione verso l'esterno.

Il primo blocco acquisisce dati georeferenziati di vario tipo. Sono integrati diversi tipi di sensore (ambientali, elettrici, meteo), uniformando i dati ricevuti e inoltrandoli al context broker ORION – punto di snodo per i flussi di dati acquisiti. Inoltre, sono integrati

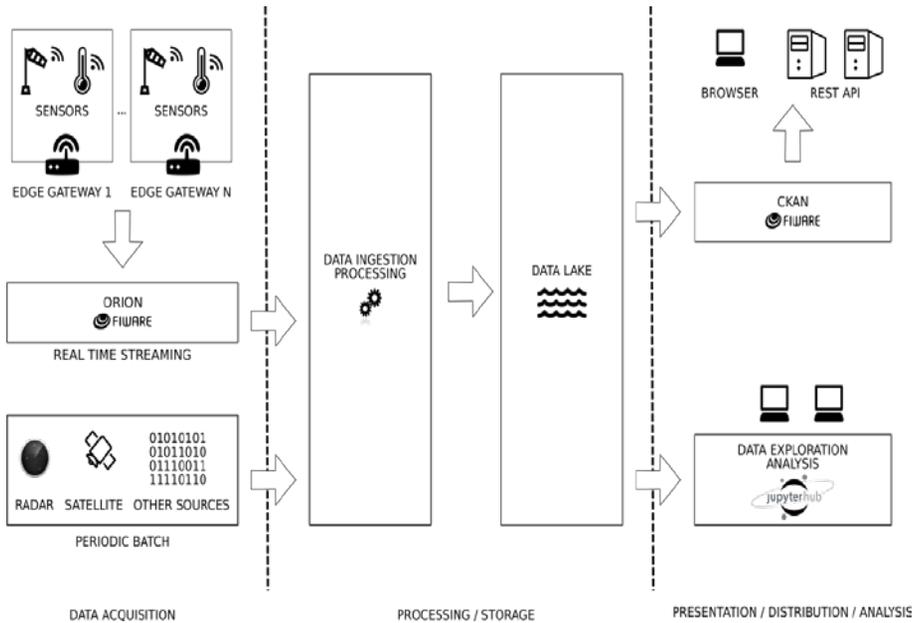


Fig. 1  
L'architettura generale di TDM

vari workflow di ingestione che a intervalli regolari acquisiscono dati georeferenziati da sorgenti come i satelliti Sentinel, radar meteorologici o dati sul traffico urbano. Il secondo blocco accumula, processa e analizza questi dati assieme a quelli provenienti da altre sorgenti sintetiche come output di simulazioni meteo. Il terzo blocco rende accessibili i dati come OpenData.

Un'architettura scalabile di aggregazione, elaborazione e restituzione provvede all'integrazione dei dati all'interno del secondo blocco. In essa sono distinguibili tre distinti sottoblocchi.

Il primo si occupa di acquisire i dati grezzi e processarli, eseguendo elaborazioni quasi real-time di quelli provenienti da sensore e elaborazioni più lunghe (batch) per operazioni complesse quali le previsioni meteorologiche a breve periodo (nowcasting), le analisi storiche dei consumi energetici, la creazione di modelli per la simulazione e la loro visualizzazione. In tale contesto sono state sviluppate tecnologie abilitanti per la compressione e visualizzazione di dati volumetrici (Marton, 2019).

Il secondo è un data lake scalabile in cui vengono memorizzati sia i dati grezzi che i dati prodotti da simulazioni e da successive elaborazioni.

Il terzo è invece dedicato all'esplorazione dei dati, principalmente attraverso l'utilizzo di servizi JupyterHub integrati.

L'intera infrastruttura è definita in codice (infrastructure as code), pertanto risulta facilmente riproducibile e modificabile programmaticamente.

## 2. Sensoristica

L'acquisizione diffusa dei dati avviene attraverso piattaforme hardware a basso costo, facilmente reperibili in commercio e distribuite sul territorio dai cittadini. L'introduzione di un Edge Gateway (EG) – un dispositivo sviluppato nel corso del progetto – disaccoppia

l'acquisizione della misura dall'invio e permette di gestire l'eterogeneità di semantica e sintassi dei dati, convertendoli in un formato standard. L'EG permette inoltre di sopperire a limitazioni dei sensori quali ridotte risorse computazionali, di memorizzazione e rete, sicurezza. L'EG, di uso generale, è stato realizzato per essere utilizzato con infrastrutture basate su FIWARE o cloud per l'IoT diverse da TDM. I suoi principali vantaggi sono la possibilità di memorizzare, filtrare e aggregare in locale i dati provenienti da più sensori, di permettere la modulazione della quantità di dati trasmessi e la frequenza di invio verso il cloud, di pre-processare i dati, e di gestire la privacy.

### **3. Prime applicazioni verticali**

L'infrastruttura e la sensoristica sono usate nell'ambito del progetto per diverse applicazioni verticali. Descriviamo qui le applicazioni relative al contesto meteo-ambientale ed energetico.

#### **3.1 Applicazioni meteo-ambientali**

Obiettivo generale di questa applicazione è studiare, sviluppare e sperimentare metodologie di nowcasting per scopi di gestione del rischio, e di procedere all'intervalidazione di modelli e dati da sensoristica diffusa.

A tal fine, sensori per le principali variabili ambientali (temperatura, pressione, umidità, precipitazione, vento e polveri sottili) sono stati realizzati e sono in via di distribuzione ai cittadini dell'area metropolitana. Queste misure saranno combinate con quelle fornite da un radar meteorologico posizionato presso il Dipartimento di Idraulica di UNICA (Farris, 2018) e con le informazioni di sensori pre-esistenti. Poiché uno degli obiettivi specifici è quello di fornire previsioni del verificarsi di eventi intensi di precipitazione e di ondate di calore nell'area metropolitana, una catena modellistica è in via di realizzazione sia per il nowcasting che per l'intervalidazione dei dati della rete dei sensori. La catena meteorologica ad area limitata (LAM) allo stato dell'arte è stata già implementata (modelli: BOLAM, MOLOCH e WRF); utilizza come condizioni al contorno ed iniziali i dati GFS e fornisce una dettagliata previsione locale a breve scadenza. L'obiettivo successivo è quello di inglobare i dati dei sensori locali e del radar nel sistema di previsione per fornire nowcasting di maggiore dettaglio e affidabilità. I dati della sensoristica diffusa saranno resi accessibili come Open Data, opportunamente validati mediante confronto incrociato con i dati simulati utilizzando la piattaforma del progetto.

#### **3.2 Consapevolezza energetica**

TDM intende incrementare la consapevolezza energetica della popolazione, attraverso il monitoraggio di consumi e risorse, per favorire la creazione di reti intelligenti di distribuzione e autoconsumo di energia. Il monitoraggio continuo e aperto della produzione e dei consumi energetici avverrà attraverso una mappatura della potenzialità di produzione energetica da fonte rinnovabile, della produzione e dei consumi energetici cittadini, sfruttando l'infrastruttura big data e la sensoristica di TDM. Un'applicazione specifica riguarda la previsione della produzione energetica da rinnovabile, integrando le

informazioni sulle variabili ambientali e sulla produzione di impianti campione ottenute mediante sensoristica diffusa con le previsioni ottenute dai modelli meteorologici sviluppati nel progetto (Massidda-1, 2018).

Riguardo ai consumi, l'obiettivo è l'analisi puntuale dei consumi elettrici degli impianti e la loro previsione, utilizzando la rete di sensori energetici e le elaborazioni dei modelli meteorologici (Massidda-2, 2018). Il monitoraggio dei consumi elettrici e della generazione da fonti rinnovabili in abitazioni private ed in edifici pubblici avverrà attraverso l'utilizzo di sistemi Open Hardware ed Open Software . I dati messi a disposizione della cittadinanza, accumulati attraverso i sensori, saranno restituiti come Open Data. Infine, si svilupperà uno strumento di pianificazione, dedicato alla valutazione dei benefici tecnico-economici associati alla gestione di micro-reti elettriche per ottimizzare gli investimenti nel settore energetico

## Conclusioni

Abbiamo illustrato l'architettura generale e le prime applicazioni del progetto TDM. Allo stato attuale, il sistema e le applicazioni sono funzionanti. Nel corso del prossimo anno sarà effettuata la distribuzione a scala urbana dei sensori, con il coinvolgimento sia di enti pubblici che di privati cittadini. TDM è finanziato dalla Regione Autonoma della Sardegna su fondi POR FESR 2014-2020, Azione 1.2.2.

In memoria del Dr. Gianluigi Zanetti, il cui esempio continua ad ispirarci.

## Riferimenti bibliografici

S. Farris, A. Seoni, D. Ruggiu, S. Bertoldo, G. Perona, M. Allegretti, M. Marrocu, M. Badas, F. Viola, & R. Deidda, First analyses of rainfall patterns retrieved by a newly installed x-band radar over the metropolitan area of Cagliari (Sardinia, Italy), 11th Int. Workshop on Precipitation in Urban Areas, Posters, 2018.

F. Marton, M. Agus, & E. Gobbetti, A framework for GPU-accelerated exploration of massive time-varying rectilinear scalar volumes, *Computer Graphics Forum* 38(3): 53-66, 2019.

L. Massidda & M. Marrocu, Quantile regression post-processing of weather forecast for short-term solar power probabilistic forecasting, *Energies* 11(7): 1763, 2018.

L. Massidda & M. Marrocu, Smart meter forecasting from one minute to one year horizons, *Energies*, 11(12): 3520, 2018.

<http://www.tdm-project.it>

<https://www.fiware.org>

<http://cfconventions.org>

<https://github.com/tdm-project/tdm-edge/tree/develop>

<http://meteoradar.unica.it>

<http://demo.tdm-project.it/eventi>

IoTaWatt, <http://iotawatt.com>

## **Autori**

**Fabio Bettio** - [fabio.bettio@crs4.it](mailto:fabio.bettio@crs4.it)

Fabio Bettio è Senior Technologist e responsabile del Visual Computing Lab al CRS4.

**Giovanni Busonera** - [giovanni.busonera@crs4.it](mailto:giovanni.busonera@crs4.it)

Giovanni Busonera fa parte del settore Data-intensive Computing del CRS4, dove si occupa di analisi dei dati e di Machine Learning con particolare riferimento alle applicazioni biomediche.

**Marco Cogoni** - [marco.cogoni@crs4.it](mailto:marco.cogoni@crs4.it)

Marco Cogoni fa parte del settore Data-intensive Computing del CRS4, dove si occupa di modellazione e simulazione di sistemi complessi.

**Roberto Deidda** - [rdeidda@unica.it](mailto:rdeidda@unica.it)

Roberto Deidda, Professore Ordinario di Costruzioni Idrauliche, Marittime ed Idrologia, Università di Cagliari.

**Mauro Del Rio** - [mauro.delrio@crs4.it](mailto:mauro.delrio@crs4.it)

Mauro Del Rio fa parte del settore Data-intensive Computing del CRS4, dove si occupa di Smart Cities e computazione distribuita di dati clinici.

**Massimo Gaggero** - [massimo.gaggero@crs4.it](mailto:massimo.gaggero@crs4.it)

Massimo Gaggero è Senior Technologist presso il settore Data-intensive Computing, Programma Distributed Computing del CRS4.

**Enrico Gobbetti** - [enrico.gobbetti@crs4.it](mailto:enrico.gobbetti@crs4.it)

Enrico Gobbetti è dirigente del settore Visual Computing del CRS4.

**Simone Leo** - [simone.leo@crs4.it](mailto:simone.leo@crs4.it)

Simone Leo lavora come ricercatore presso il Settore Data-intensive Computing, programma Distributed Computing del CRS4.

**Simone Manca** - [simone.manca@crs4.it](mailto:simone.manca@crs4.it)

Simone Manca lavora come Software Engineer presso il CRS4 dal 2000. Attualmente si occupa di Spatial Data Infrastructure e Applicazioni Ambientali.

**Marino Marrocu** - [marino.marrocu@crs4.it](mailto:marino.marrocu@crs4.it)

Ricercatore senior, responsabile del programma di ricerca Environmental Sciences del CRS4.

**Luca Massidda** - [luca.massidda@crs4.it](mailto:luca.massidda@crs4.it)

Ricercatore senior, responsabile del programma di ricerca Smart Energy Systems del CRS4.

**Fabio Marton** - [fabio.marton@crs4.it](mailto:fabio.marton@crs4.it)

Fabio Marton dal 2002 è ricercatore nel settore Visual Computing del CRS4, dove attualmente

dirige il programma Visual and Geometric Computing.

**Marco Enrico Piras** - [marcoenrico.piras@crs4.it](mailto:marcoenrico.piras@crs4.it)

Marco Enrico Piras lavora come tecnologo presso il Settore Data-intensive Computing, programma Distributed Computing del CRS4.

**Luca Pireddu** - [pireddu@crs4.it](mailto:pireddu@crs4.it)

Ricercatore col settore Data-intensive Computing del CRS4, attualmente lavora su problemi legati al deployment e alla scalabilità di applicazioni su IaaS.

**Gabriella Pusceddu** - [gabriella.pusceddu@crs4.it](mailto:gabriella.pusceddu@crs4.it)

È ricercatore expert al CRS4 e fa parte del Programma Scienze Ambientali del Settore Energia e Ambiente. I suoi interessi sono rivolti all'uso di modelli numerici ad area limitata per la previsione di eventi meteorologici estremi e al downscaling di scenari climatici.

**Alessandro Seoni** - [aseoni@unica.it](mailto:aseoni@unica.it)

Alessandro Seoni, Tecnico Laureato, Università di Cagliari.

**Gianluigi Zanetti** - [gianluigi.zanetti@crs4.it](mailto:gianluigi.zanetti@crs4.it)

Gianluigi Zanetti è direttore del settore Data-intensive Computing del CRS4.

## L'entrepreneurship education ed il progetto One Stop Shop

Gianluigi De Gennaro, Annamaria Demarinis Loiotile, Annalisa Turi

Università di Bari Aldo Moro, Innovation Creativity Center

**Abstract.** Il mondo (alias la società), soprattutto grazie agli sviluppi tecnologici, cambia continuamente e con esso cambiano il lavoro, la cultura e l'economia. L'innovazione risulta l'elemento chiave ed i processi di trasferimento tecnologico e di open innovation diventano fondamentali non solo per la valorizzazione delle invenzioni, ma soprattutto per la competitività di imprese e territori in un contesto globalizzato. L'Università di Bari è best practice grazie ad un progetto che coniuga engagement, exchange ed enterprise

**Keywords.** Innovazione, trasferimento tecnologico, entrepreneurship, startup, open innovation

Oggi, in quella che si potrebbe definire “società dell'innovazione aperta e partecipata”, l'innovazione risulta l'elemento chiave dei processi di trasformazione socio-economici che porta ad un ribaltamento – in senso positivo – da un uso meramente strumentale delle tecnologie ad una loro valorizzazione sociale ponendo al centro il ruolo della persona nel nuovo scenario di Industria 4.0.

Così “nell'evoluzione economica attuale, lo sviluppo di idee e progetti creativi assume un valore sempre più rilevante grazie alla loro capacità di generare ricchezza e “proprietà intellettuale” (brevetti, diritti d'autore, marchi di fabbrica, design registrato) e di sostenere così anche lo sviluppo economico di un Paese. L'ingresso di nuovi imprenditori e di nuove idee è, infatti, veicolo di innovazione di prodotti e processi e alimenta quella interazione tra persone, idee e capitali che consente lo sviluppo e la creazione di spazi occupazionali, l'avvio di cicli virtuosi di accumulazione di conoscenza tecnologica e organizzativa, nonché la possibilità di creazione di nuove imprese” .

In questo nuovo approccio, i flussi della ricerca pubblica e – di conseguenza – del trasferimento tecnologico diventano contenitori, e al tempo stesso contenuti, per generare nuova conoscenza e valorizzandola sotto varie forme, tra cui proprio l'open innovation. Con il termine Trasferimento Tecnologico (TT) si intende definire il processo di valorizzazione della ricerca scientifica in prodotti e servizi trasferibili, ergo che possono essere commercializzati. E' lapalissiano che le Università e gli Enti pubblici di ricerca assumono un ruolo cruciale; la letteratura economica, infatti, oltre ad aver evidenziato il contributo della ricerca universitaria all'innovazione industriale, ha anche dimostrato come le interazioni tra il mondo accademico e quello imprenditoriale abbiano effetti positivi per entrambi.

L'imprenditorialità sta diventando un'opzione di occupazione sempre più attraente, non solo tra persone altamente qualificate e con esperienza, ma anche tra i più giovani co-

me i laureati . Ciò è dimostrato dal crescente tasso di creazione di nuove imprese da parte degli studenti e dalla crescente domanda di programmi formativi riguardanti l'imprenditorialità e di strutture di sostegno all'imprenditorialità nelle università di tutto il mondo .

L'imprenditorialità è il motore della crescita economica; contribuisce alla creazione di nuove imprese e di posti di lavoro, allo sviluppo di nuove capacità e competenze. Anche in Italia un numero crescente di studenti considera l'imprenditorialità un'opzione di carriera realistica, con numerosi esempi di studenti che hanno fondato le proprie iniziative imprenditoriali durante gli studi universitari o subito dopo la laurea .

In Italia, le Università hanno sempre tradizionalmente fornito un contributo fondamentale allo sviluppo sociale ed economico del proprio territorio; ma oggi sono sempre più spinte ad essere protagoniste nei processi di valorizzazione della ricerca accademica. Coniugare scienza e mercato non è però un processo immediato: è necessario sviluppare nei ricercatori un orientamento all'innovazione e all'imprenditorialità che si affianchi alla ricerca indotta da curiosità e approccio esclusivamente scientifico.

“L'importanza che la valorizzazione della ricerca va assumendo all'interno dei Sistemi Universitari Pubblici e nei Centri di Ricerca impone sempre più una maggiore attenzione non solo ai team e agli attori della ricerca, ma anche a quelle modalità organizzative capaci di fare la differenza” .

In particolare, negli ultimi anni, l'Università degli Studi di Bari Aldo Moro, soprattutto grazie al Progetto “ONE STOP SHOP for Engagement, Exchange, Enterprise” , ha implementato un modello di trasferimento tecnologico (ispirato a quello delle Università israeliane) che non solo costituisce uno degli assi portanti della cosiddetta “terza missione”, ma anche l'ancora per agganciare il futuro. Il Progetto “One Stop Shop” è proprio finalizzato ad offrire ai ricercatori una serie di servizi che si rendono necessari in quella che è ormai nota come “knowledge economy”, adottando un modello di TT che bene si adatta alle peculiarità ed ai bisogni del territorio. “One Stop Shop” rappresenta nell'Università di Bari la cerniera nelle relazioni tra ricerca pubblica e mercato/industria, l'hub verso la domanda, l'ufficio dell'account, il marketing dell'offerta, la valorizzazione dei risultati delle ricerche in un processo continuo di entrepreneurial discovery, la porta di ingresso per le richieste di matching di competenze ed esigenze di innovazione industriale, il coordinamento degli spin-off ed il fronte più vivace della creatività.

Le Università stanno pertanto provando a diventare agenti di innovazione regionale nel senso più strettamente schumpeteriano , trasformandosi in ‘Entrepreneurial Universities’ dove le relazioni con gli stakeholderes, il territorio, le imprese sono aspetti vitali per la creazione di una strategia innovativa di apprendimento per gli studenti .

L'Entrepreneurship Education è una strategia fondamentale per creare le condizioni adeguate alla creazione di una innovazione knowledge-based; pertanto, alle comunità accademiche è richiesto un particolare sforzo per incoraggiare e formare una nuova generazione di imprenditori innovativi che sono il reale motore per una crescita economica duratura, sostenibile ed inclusiva.

Relativamente a questo tema, l'Università di Bari ha avviato ed anima continuamente un Contamination Lab, denominato BaLab, laboratorio avente l'obiettivo di favorire la

contaminazione di idee imprenditoriali innovative. Nel continuum del processo di contaminazione, cadenzato da riunioni strutturate di analisi condivisa dei singoli progetti d'impresa, si inseriscono attività di team building, co-progettazione e co-costruzione (a partire dagli spazi stessi del BaLab), eventi formativi ed incontri con le opportunità. Un contesto interdisciplinare dove si sperimentano nuovi modelli formativi sempre più orientati ai bisogni di un'utenza che rivendica protagonismo nel percorso di sviluppo delle competenze (formazione on-demand e laboratori dal basso).

Nell'ambito del BaLab, sono nate e si sono sviluppate numerose startup; diversi gruppi hanno vinto competition nazionali ed internazionali; a titolo esemplificativo si riportano due casi di successo.

Nata a Bari nel 2015, Mindesk si è fatta spazio nel mondo della realtà virtuale grazie ad una piattaforma di modellazione che permette a molteplici utenti di creare o revisionare oggetti tridimensionali in tempo reale. Il software è disponibile come plug-in per alcuni degli ambienti Cad più diffusi (McNeel Rhinoceros, Grasshopper e Dassault SolidWorks) nel campo dell'ingegneria, dell'industria navale, del design industriale e delle arti creative. Nextome è un sistema di navigazione e localizzazione che opera in ambienti chiusi permettendo all'utente di conoscere con precisione la propria posizione all'interno di edifici, pianificare e seguire percorsi e godere di una serie di servizi ad essa connessi. Grazie ad un vincente matching tra il Dipartimento di Biologia dell'Università di Bari e due startup (oltre Nextome, anche Energy By OSCAR), si è dato avvio al progetto "CLEAN aiR in Hospital", prototipo pilota per valutare la qualità dell'aria nelle strutture ospedaliere, installato presso l'Istituto Tumori Giovanni Paolo II di Bari.

Per l'Università sostenere i giovani innovatori nel costruire il proprio futuro vuol dire anche sostenere il futuro del territorio con un ruolo da protagonista.

## Riferimenti bibliografici

M. Costa, Il valore generativo del lavoro nei contesti di Open innovation, in Rivista internazionale di Scienze dell'educazione e della formazione, vol. 9 - n. 3, 2011.

L. Petretto, 2008, Imprenditore ed Università nello start-up di impresa. Ruoli e relazioni critiche, Firenze University Press.

V. E. Mansfield, 1998, Academic research and industrial innovation: An update of empirical findings, University of Pennsylvania.

V. R. Coombs – A. Richards – P.P. Saviotti - V. Walsh (a cura di), 1996, Technological Collaboration: The Dynamics of Cooperation in Industrial Innovation, Edward Elgar.

R. Fini - A. Meoli - M. Sobrero - S. Ghiselli - F. Ferrante, 2016, Student Entrepreneurship: Demographics, Competences and Obstacles.

OECD. 2015a. Entrepreneurship at Glance 2015. OECD Publishing, Paris.

Å. Lindholm Dahlstrand – E. Berggren, 2010, Linking innovation and entrepreneurship in higher education: a study of Swedish schools of entrepreneurship. In: R., Oakey, A., Groen, G., Cook, & P., Van der Sijde, (Eds.), New Technology-Based Firms in the New Millennium, vol. 8, pp. 35–50. Emerald Group Publishing.

D. Dell'Anno, 2010, La conoscenza dall'università all'impresa. Processi di trasferimento tecnologico e sviluppo locale, Carocci.

XIII Rapporto Netval sulla Valorizzazione della Ricerca Pubblica Italiana, 2016.

Bando del Ministero dello Sviluppo Economico per il Potenziamento degli UTT per una loro maggiore focalizzazione degli UTT sulla protezione e trasferimento dei titoli di proprietà industriale relativi a specifici settori produttivi.

V. J. A. Schumpeter, 1971, Teoria dello sviluppo economico. Ricerca sul profitto, il capitale, il credito, l'interesse e il ciclo economico.

P.S. Vincett - S. Farlow, 2008, 'Start-a-Business': An experiment in education through entrepreneurship, in Journal of Small Business and Enterprise Development, 15(2), pp. 274-288

## Autori



**Gianluigi de Gennaro** - [gianluigi.degennaro@uniba.it](mailto:gianluigi.degennaro@uniba.it)

Ricercatore di Chimica dell'Ambiente presso il Dipartimento di Biologia dell'Università degli Studi di Bari Aldo Moro. Componente della Commissione nazionale VIA-VAS. Responsabile Scientifico del BaLab, il contamination lab dell'Università di Bari. Già Delegato del Rettore allo Sviluppo della Creatività; Presidente del Centro di Eccellenza "Per l'Innovazione e la Creatività" e Referente dell'Industrial Liaison Office e del Progetto "One Stop Shop", cofinanziato dal Ministero dello Sviluppo Economico. È fondatore dello Spin-Off accademico LEnviro5.

**Annamaria Demarinis Loiotile** - [annamaria.demarinis@uniba.it](mailto:annamaria.demarinis@uniba.it)

Laurea con lode, nel 2006, in Chimica; Master in "Manager della ricerca e Tecnico commerciale per il mercato dell'innovazione" e Corso di Alta Formazione in "Comunicazione della ricerca". Lavora, dal 2007, all'Università di Bari, occupandosi di progetti di ricerca riguardanti la chimica dell'ambiente e la sostenibilità ambientale. Oggi, nell'ambito del Centro di Eccellenza "Per l'Innovazione e la Creatività" di UNIBA, con il ruolo di Research Manager, si occupa di progetti di ricerca, innovazione e trasferimento tecnologico, rapporti con le imprese ed il territorio.



**Annalisa Turi** - [annalisa.turi@uniba.it](mailto:annalisa.turi@uniba.it)

Laurea in Giurisprudenza; Master in "Fiscalità d'impresa" e Corso di Alta Formazione sulle Tecnologie abilitanti per la formazione "EDOC@Work". Avvocato; si definisce un giurista atipico con propensione per l'innovazione e la creazione di impresa.

Dottorando di ricerca in Diritto Tributario, con particolare attenzione per la fiscalità dell'innovazione. Numerose pubblicazioni su startup e fiscalità promozionale. Knowledge Transfer Manager presso il Centro di Eccellenza "Per l'Innovazione e la Creatività" dell'Università di Bari Aldo Moro.

# Un caso riuscito di contaminazione scientifica tra ambienti diversi

Giuseppe Di Persio

Istituto Nazionale di Astrofisica

**Abstract.** Viene brevemente descritta la genesi del formato FITS, nato in ambiente astronomico, la sua evoluzione e come sia divenuto parte integrante del progetto di digitalizzazione dei manoscritti della Biblioteca Apostolica Vaticana.

**Keywords.** Conservazione Digitale, Biblioteche, Astronomia

## Introduzione

Si spiegherà come un formato usato solo in Astronomia sia approdato e sia stato usato dalla Biblioteca Vaticana nel suo progetto di conservazione digitale.

### 1. Nasce un problema, nasce il FITS

Negli anni 70 ci fu un grande sviluppo dell'Astronomia e dell'Astrofisica. I sempre più numerosi satelliti in orbita, i telescopi e i radiotelescopi sulla Terra produssero una enorme (per l'epoca) quantità di dati. Questo fece sorgere alcuni problemi, il principale dei quali era la pochissima standardizzazione nei formati di salvataggio, fattore che ostacolava molto la diffusione del sapere e lo studio condiviso.

La comunità scientifica astronomica raccolse la sfida, studiò il problema e nel 1979 furono rese pubbliche le specifiche di un nuovo formato di salvataggio dei dati: era nato il FITS (Flexible Image Transport System), standard dal 1981 e che divenne ben presto il formato più usato per la memorizzazione di immagini e dati in Astronomia. Fino al 1987 questo formato era gestito direttamente dalla NASA tramite un suo dipartimento (il NASA Office of Standard and Technology), poi la gestione passò allo IAU (International Astronomical Union), la più importante associazione di scienziati del settore.

### 2. Cambiare rimanendo se stessi

Il FITS è usatissimo ancora oggi a quasi quarant'anni dalla sua nascita, un periodo che in informatica equivale a un'era geologica. I motivi di questo successo sono da ricercarsi nel fatto che è stato creato tenendo ben presente il problema della conservazione digitale a lungo termine, quindi è semplice ed efficiente.

Le sue caratteristiche principali sono:

- è esente da royalty e di libero uso,

- tutte le informazioni necessarie a leggere, capire e rappresentare i dati contenuti sono facilmente leggibili e accessibili,
- la sua ampia diffusione e uso sono garanzia di longevità,
- è flessibile, potendo salvare molti tipi di dati con una grande quantità di metadati, ma robusto relativamente alla possibile perdita di informazioni,
- esiste un'ampia documentazione sul formato tutta di pubblico dominio,
- è completamente indipendente dal sistema operativo in uso,
- è personalizzabile senza perdere l'aderenza allo standard,
- gestisce senza problemi file di grandi dimensioni.

Un file FITS è composto da due elementi distinti, che si possono ripetere più volte. Il primo elemento è l'HEADER, formato da testo ASCII visualizzabile facilmente in ogni ambiente operativo, mentre il secondo elemento è composto dai dati in formato binario. Nell'HEADER ci sono delle "Keyword" che definiscono dei valori per una serie di parametri: alcuni (pochi) obbligatori, altri facoltativi che servono ad inserire informazioni aggiuntive. Questo schema HEADER-DATA si può ripetere più volte e permette di memorizzare nello stesso file più immagini o dati correlati.

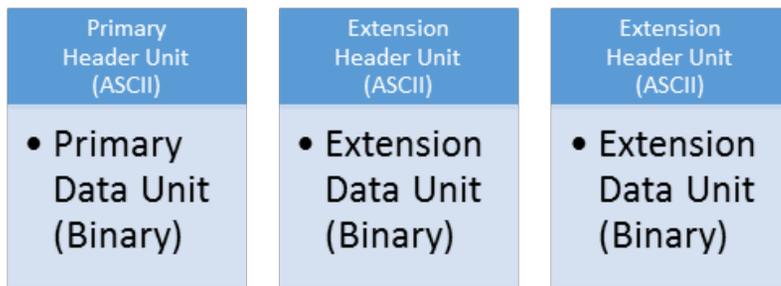


Fig. 1  
Struttura di un file FITS

Esempio di HEADER di un file FITS:

Keyword Obbligatorie:

SIMPLE = T / Il file è conforme allo standard

BITPIX = 16 / Numero di bit per pixel

NAXIS = 2 / Numero degli assi

NAXIS1 = 1024 / Risoluzione dell'asse 1

NAXIS2 = 1024 / Risoluzione dell'asse 2

Keyword facoltative:

AUTHOR = ' JOHN SMITH'

DATE = '05/06/1984'

OBJECT = ' True photos of aliens on Mars'

ORIGIN = ' ASA secret archive'

DATAMAX = 1.246536493E+05 /MAX PIXEL VALUE

DATAMIN = 1.483317930E+01 /MIN PIXEL VALUE

COMMENT = ' Esempio di un possibile HEADER'

HISTORY -----

HISTORY Permission is granted for publication and reproduction of this  
HISTORY material for scholarly, educational, and private non-commercial  
HISTORY use.

HISTORY NRAO  
HISTORY 520 Edgemont Road  
HISTORY Charlottesville, VA 22903-2475  
HISTORY -----  
END

Attualmente gran parte dei dati astronomici ed astrofisici sono memorizzati in FITS. Lo IAU cura l'evoluzione del formato, giunto alla versione 4.0 nel luglio 2016, in modo da mantenerlo sempre in grado di rispondere alle nuove esigenze della scienza, ed è prevista una articolata struttura basata su comitati locali per raccogliere e discutere le richieste di cambiamento che arrivano dalla comunità scientifica internazionale.

Ogni cambiamento introdotto ha arricchito il formato di nuove funzionalità ma ha sempre mantenuto la piena compatibilità con le precedenti versioni. Il motto ufficiale del formato: "Once FITS, always FITS", finora non è mai venuto meno!

### 3. Il FITS e la Biblioteca Apostolica Vaticana (BAV)

La BAV, tra le più importanti biblioteche al mondo, da oltre 500 anni istituzionalmente svolge un ruolo non solo di diffusione ma anche di conservazione, protezione e restauro del patrimonio librario custodito, soggetto a continua usura anche tramite il semplice contatto con la pelle umana.

Per conservare i suoi manoscritti fin dal 2007 fu avviato uno studio di fattibilità per comprendere come le tecnologie informatiche potessero aiutare a risolvere il problema, e la quasi scontata conclusione fu quella di digitalizzare il suo intero patrimonio librario. Naturalmente la digitalizzazione doveva essere ad altissima risoluzione, con una fedeltà assoluta dei colori e con l'obiettivo, ambizioso, di rendere fruibili i files dei manoscritti almeno per i prossimi 100 anni.

Se si pensa alla rapida obsolescenza di quasi tutto quello che ha a che fare con l'informatica si capisce come il lavoro non sia stato per niente facile. I pilastri del progetto di digitalizzazione dovevano essere il formato di memorizzazione e una serie di procedure volte ad evitare l'obsolescenza tecnologica dei sistemi informatici.

Mentre per il secondo punto sono state stabilite alcune stringenti linee guida e presi accordi con le aziende fornitrici dell'hardware per periodici aggiornamenti almeno biennali dei sistemi, per quanto riguarda il primo punto sono stati fatti attenti studi comparativi tra i vari formati esistenti e presa in considerazione la possibilità di crearne uno nuovo.

Purtroppo, nessuno dei formati tra quelli maggiormente diffusi in ambito fotografico aveva i requisiti necessari. Neanche quello finora più usato, il TIFF, perché:

- è gestito da una compagnia commerciale privata, e non è escluso che in futuro siano esigibili pagamenti di royalties
- non gestisce file maggiori di 4 GB mentre già esistono scanner che producono file più

grandi

- non lavora a 64 bit e non permette di salvare tutta una serie di importanti informazioni legate alle immagini, come quelle storiche o bibliografiche.

Alle prese con questo problema e venuta a conoscenza del FITS la BAV ne intuì presto le potenzialità, anche se non era mai stato usato al di fuori dell'ambito astronomico, e iniziò una serie di test.

Vista la provenienza del FITS la Biblioteca si avvalse fin dall'inizio della collaborazione di esperti dell'Istituto Nazionale di Astrofisica, e iniziò così una sinergia tra settori apparentemente diversi accumulati dai problemi della Long Term Data Preservation.

Questa collaborazione si è rivelata estremamente proficua e ha interessato anche altre importanti istituzioni del settore spaziale come lo IAU, che ha attivamente collaborato per arricchire il FITS con delle keyword specifiche per il settore della conservazione digitale, o l'ESA (European Space Agency), da sempre interessata alle problematiche della LTDP, che ha firmato con la BAV degli accordi per condurre studi congiunti nel settore.

Dopo una serie positiva di test la BAV ha fatto partire il progetto di digitalizzazione dei manoscritti.

I numeri del progetto:

- • 80.000 manoscritti
- • 150.000 stampe, mappe e disegni
- • 9.000 incunaboli
- • +130.000 unità archivistiche
- • Tutti scansionati ad altissima risoluzione
- • Oltre 40 PB di dati originali

Il progetto, che ha suscitato l'interesse di altre istituzioni del settore, è basato sull'uso di FITS per salvare le immagini e su XML per tutte le informazioni bibliografiche o di altra natura non gestibili da quel formato (il FITS salva dati in formato ASCII e/o binario, e non permette di salvare testo in caratteri non ASCII). Al momento è stato digitalizzato poco più del 20% del patrimonio librario della Biblioteca.

## Conclusioni

La digitalizzazione dei manoscritti è stata non un punto di arrivo ma di partenza per una serie di ulteriori interessanti sviluppi.

In generale solo una frazione del patrimonio librario conservato nelle grandi Biblioteche è attentamente studiata, essenzialmente a causa della delicatezza dei manoscritti e del fatto che devono essere studiati di persona. La digitalizzazione di altissima qualità in corso permetterà agli studiosi di analizzare i testi senza dover spostare o toccare i manoscritti, operazione facilitata anche dalla partecipazione della BAV come socio fondatore al Consorzio IIF e quindi dalla possibilità di usare questo interessante standard per la visualizzazione in alta definizione degli elementi digitalizzati.

Inoltre si potranno usare tecniche automatiche di riconoscimento del testo o di immagini che permetteranno indagini e studi su scala mai tentata prima.

Un ultimo ma credo significativo risultato è che la BAV ha reso disponibili i manoscritti digitalizzati, in una versione a più bassa risoluzione, in un apposito sito che permette a chiunque di ammirare gratuitamente questi capolavori.

## Riferimenti bibliografici

Siti della BAV dedicati al Progetto di Digitalizzazione:

<https://www.vaticanlibrary.va/home.php?pag=digitalizzazione&ling=eng>

<https://digi.vatlib.it>

Home page del FITS

[https://fits.gsfc.nasa.gov/fits\\_home.html](https://fits.gsfc.nasa.gov/fits_home.html)

## Autore



**Giuseppe Di Persio** - [giuseppe.dipersio@inaf.it](mailto:giuseppe.dipersio@inaf.it)

Nasce a Roma nel 1964. Si diploma in elettronica, ma l'informatica diventa presto la sua passione. Dal 1984 lavora al CNR presso diversi istituti, per poi confluire nel 2003 all'INAF a Roma. Nel 2002 si laurea in Informatica.

È Computer Scientist presso lo Space Science Data Center dell'ASI e fa parte del team software di diversi satelliti. Si occupa di programmazione, acquisizione e analisi dati, progettazione e gestione di data center scientifici, sicurezza e reti. Negli ultimi anni anche di Long Term Data Preservation, collaborando con la Biblioteca Vaticana nel suo progetto di digitalizzazione tramite l'uso di standard astronomici.

# Managing research data at the University level: the experience of the University of Milan

Paola Galimberti<sup>1</sup>, Susanna Mornati<sup>2</sup>

<sup>1</sup>University of Milan, <sup>2</sup>4Science s.r.l

**Abstract.** The University of Milan is the largest university in Lombardy and one of the largest in Italy, and the only Italian university to be part of the League of European Research Universities (LERU). As an observer on the international scene, it has witnessed a series of activities on research data at European level and the lukewarm transposition, at national level, of policies decidedly oriented towards the opening up of research processes. Among the recommendations concerning Open Science issued by the European Commission [1], the management of FAIR data [2] is one of the main pillars identified by LERU in the roadmap for universities [3]. Moving from observer to protagonist was a decisive step for the University of Milan therefore, confirming the responsibility of leading role among Italian universities assumed with the participation in the LERU. Little has been said about research data so far in Italy, always and only among technicians, little has been asked and little has been built. There is no central data infrastructure in Italy, although initiatives are being taken at group level. We are also of the opinion that the individual institutions should have a responsibility in the collection, management, care, conservation and transmission of the heritage of knowledge, a responsibility that has been instanced with the adoption of institutional repositories for scientific literature, and now requires a step forward, facilitated by the availability of ad hoc technologies, open and simple and inexpensive to adopt, to manage the research data produced at the university level. At the University of Milan, we tried to intercept the needs of users even before they were expressed. Researchers in the departments were interviewed to understand what the real needs were, but above all to try to understand the degree of awareness with regard to the issue of data storage and retention, data responsibility and the main critical issues in their management. The interviews revealed a varied picture with some points in common: need for technological support, need for training, need for legal support. With regard to technological support, we have tried to identify and make available to researchers a tool that could meet the requirements internationally, that was open source, that could also interface with other tools of the university. The choice fell on Dataverse, a free open source platform created within the University of Harvard and adopted by prestigious institutions around the world [4]. Following a one-year pilot project at the university, the production phase was completed at the beginning of 2019. The installation of Dataverse has been outsourced to 4Science, the only Italian company that collaborates with the international community of Dataverse and offers services in the cloud. The article explains how the choice of Dataverse came about and why it was decided to use an instrument other than AIR (the University's Institutional Repository), what the critical points were, how it is planned to proceed with the promotion of this instrument currently used by two departments, what integrations are desirable in the immediate future and what are the benefits expected from this new approach

**Keywords.** RDM, data preservation, data curation, data stewardship, Dataverse

## 1. Anticipate user needs

If at a global level there is a great deal of attention on research data and on the possibility of managing, archiving and preserving them, at a national and institutional level in Italy it is

not yet clear what the needs of researchers might be, and therefore little has been achieved.

It is in an attitude of listening to the demands at international level and, above all, to the priorities of the LERU of which the University is a member, that an attempt has been made to understand and anticipate what the requests of researchers would have been once the requests of the research funding bodies had been accepted.

In order to understand what data was produced in the different disciplinary areas, how many, where it was stored and with what rules, how it was stored, who could access it and why, who was responsible for it, a series of interviews were carried out with professors and researchers from the different departments, the most evident outcome of which was the awareness that there was no management of the data, that there was no concern about its conservation, that the legal and ethical issues related to the data were very unclear and that its publication was in many cases completely excluded.

## **2. A Policy for research data management**

The first step was therefore to develop a policy on the management of research data that provided for precise responsibilities on the data (by the university and by the Principal Investigators), which required that the management according to FAIR principles was the standard management, and defined a period of data retention. The definition of what the university means by research data and the subjects who are required to comply with the policy has been fundamental. In addition to requiring that data be managed in accordance with FAIR principles, the University has also undertaken to find and offer its researchers a suitable tool. The policy was submitted to and approved by the Academic Senate.

## **3. The search for a suitable instrument**

Interviews showed that the researchers' needs for data management were clear, among them:

- have an easy-to-use tool
- that guarantees always and in any case the accessibility of the data
- which could also be accessed by colleagues from outside the university
- that complies with the FAIR principles
- that could be harvested by Open AIRE
- that could contain different versions of the datasets
- that attributed a DOI to the datasets and make them citable
- that would allow the publication of data with ad hoc licenses

Since the institutional repository for publications, AIR (based on DSpace technology) does not meet many of these requirements, a tool was sought that could meet all these needs.

Dataverse, a software platform for RDM developed by the Harvard University, was selected for a pilot project. It was chosen for several reasons that were assessed during the pilot and confirmed the choice:

- it is free and open source, providing a reliable and sustainable tool to support the institutional policy
- it is FAIR compliant, guaranteeing appropriate participation and high visibility in the research ecosystem

- it supports researchers' needs, namely the availability of a reliable, accessible, and durable tool to share data in a restricted group before making them public, to store data safely and make data discoverable for future reuse, to publish them in the appropriate version when funders require it, to cite and reference them in journal papers, and so on.
- Dataverse was born to meet a set of requirements to improve the data-sharing infrastructure of the scholarly community [5]:
- Recognition: as much as it is important for researchers to be recognised as the authors of a journal publication, likewise this need is felt for data sets;
- Public distribution: the dissemination of knowledge must be carried out through channels that are as public as possible and not through private agreements;
- Authorization: authors should be able to control access to their data sets, whether through a license, a guestbook, or a combination of methods;
- Validation: data's existence and validity should be verifiable even without access to the data itself;
- Verification: data's authenticity should be preserved over time, through the possibility of versioning and freezing the data sets;
- Persistence: data sets should be available for future access for an indefinite time;
- Ease of use: the tool to manage data sets should be used by researchers themselves, not necessarily mediated by professional archivists;
- Legal protection: licenses associated to data set should define responsibilities related to authorship, usage, etc.

#### **4. Training and experimentation**

Once the choice of the instrument had been made, it was proposed to the various departments, selecting in particular a department that would act as a tester for the different functions.

The choice fell on a department that had included the management of research data among the objectives of the three-year plan and that willingly carried out the experimentation.

At the end of the trial, which confirmed the choice of the tool[6], advertising began in the departments, both through a guide to the use of Dataverse that was distributed, and through ad hoc presentations aimed at framing the theme of data management, its purposes and the benefits that derive from it.

The theme has also been linked to that of research integrity, particularly with regard to the possibility of reproducing research.

#### **5. The use of Dataverse**

Currently there are two main ways of using Dataverse: on the one hand it is used to deposit the datasets that are then indicated in the journals as material accompanying the articles submitted, on the other hand it is used as a reference tool for data management in data management plans that accompany European projects and beyond.

#### **6. An initial assessment**

One year after the start of the trial and as the awareness of the importance of data

management grows, the number of datasets stored has increased dramatically.

Most of the archive is not open at the moment, so the substantial part of the archive is not visible, even if it is accessible. There is still a lot of support work to be done, from producing guidelines and FAQs to help researchers find their way around, to setting up courses and tutorials.

At this point, however, we can say that a path has been undertaken, certainly challenging, but that will lead our researchers to a greater awareness about research processes, research reproducibility and its reuse even for purposes other than those for which it was designed.

## References

- [1] European Commission - Open Science: <https://ec.europa.eu/research/openscience/index.cfm>; last accessed 5 April 2019
- [2] Wilkinson, M. D. et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018, doi: 10.1038/sdata.2016.18
- [3] League of European Research Universities (2018) Open Science and its role in universities: A roadmap for cultural change, Advice Paper no. 24 May, <https://www.leru.org/files/LERU-AP24-Open-Science-full-paper.pdf>
- [4] The Dataverse Project, <https://dataverse.org/>, last accessed 5 April 2019
- [5] Gary King (2007) An Introduction to the Dataverse Network as an Infrastructure for Data Sharing. *Sociological Methods and Research* 36, pp. 173–199, <https://gking.harvard.edu/files/abs/dvn-abs.shtml>
- [6] <https://dataverse.unimi.it/>

## Authors

**Paola Galimberti** - [paola.galimberti@unimi.it](mailto:paola.galimberti@unimi.it)

Paola Galimberti is responsible for the supporting activities for the evaluation board, for the governing bodies and for the quality assurance. She manages the institutional archive IRIS (she is coordinator of the national working group Cineca), the platform of e-publishing OA ([riviste.unimi.it](http://riviste.unimi.it)) and the repository for research data ([dataverse.unimi.it](http://dataverse.unimi.it)). She supports the activities of the OS committee of the University and is delegate in the LERU group on Open Science issues.

She is a member of the Board of AISA (Italian Association for the Promotion of Open Science); she is a member of the editorial staff of ROARS (Return on Academic Research), member of IOSSG, (Italian Open Science Support Group) and editor for Italy and Germany of the DOAJ.

Paola Galimberti is responsible for the supporting activities for the evaluation board, for the governing bodies and for the quality assurance. She manages the institutional archive IRIS (she is coordinator of the national working group Cineca), the platform of e-publishing OA ([riviste.unimi.it](http://riviste.unimi.it)) and the repository for research data ([dataverse.unimi.it](http://dataverse.unimi.it)). She supports the activities of the OS committee of the University and is delegate in the LERU group on Open Science issues.

She is a member of the Board of AISA (Italian Association for the Promotion of Open Science); she is a member of the editorial staff of ROARS (Return on Academic Research), member of IOSSG,

(Italian Open Science Support Group) and editor for Italy and Germany of the DOAJ.

**Susanna Mornati** - [susanna.mornati@4science.it](mailto:susanna.mornati@4science.it)

Susanna Mornati is COO at 4Science, Italy. She has extensive experience in the design and implementation of information systems for research, gained in thirty years spent at the University of Milan, CERN and university consortia for ICT. With her vast expertise in the research domain, in 2015 she directed the program of implementing DSpace-CRIS (IRIS) at 67 Italian HE and research institutions and the IRIDE project for ORCID adoption at the national level in Italy. Both projects involved over 60,000 researchers and were successfully achieved in just a few months.

Susanna has gained an international reputation in the Open Science communities, participating in scientific boards and committees, and a speaker at numerous events. She is a member of the Research Data Alliance (RDA), the COAR Controlled Vocabularies Board, the DSpace Leadership and Steering Groups, the euroCRIS CRIS-IRs Task Group, the Italian Association for Open Science (AISA), the Italian Open Science Support Group (IOSSG).

# Strategia e politica di sviluppo per una infrastruttura digitale per la gestione e valorizzazione del patrimonio archivistico

Costantino Landino

ICAR - Istituto Centrale per gli Archivi

**Abstract.** La conservazione digitale può essere definita come l'interoperabilità con il futuro. È necessario che le Amministrazioni deputate a gestire il patrimonio informativo digitale modellino i loro sistemi per garantire i corretti processi di conservazione della nostra memoria nel tempo. L'ICAR sta realizzando una infrastruttura nazionale per l'erogazione di servizi per la gestione del patrimonio archivistico che sia in grado di evolvere verso la conservazione dei contenuti digitali archivistici. Agli Archivi di Stato e alle Soprintendenze Archivistiche si forniranno servizi per la gestione delle informazioni archivistiche, per la pubblicazione di inventari e digital library specializzate nella gestione del patrimonio digitale. L'accesso alla documentazione digitale avverrà con meccanismi di identificazione degli utenti in ottemperanza alle norme in materia di dati personali. Sono state tenute in considerazione esigenze di sostenibilità delle soluzioni informatiche e concettuali.

**Keywords.** Sistemi informativi, Sostenibilità, Digital Library, Conservazione digitale.

## 1. Introduzione

L'Istituto centrale per gli archivi (ICAR) è l'organismo all'interno della Direzione Generale per gli Archivi che cura l'elaborazione di norme e standard per la descrizione archivistica, per la gestione di oggetti digitali e per l'interoperabilità fra sistemi informatici. Dal 2012 cura il coordinamento tecnico-scientifico del Sistema Archivistico Nazionale (SAN) volto a integrare i sistemi informatici di ambito nazionale, regionale e locale. Gestisce l'interoperabilità tra sistemi informativi e la cooperazione con i network culturali nazionali e internazionali e pubblica i propri dati in formati Open Data e Linked Open Data.

Dal 2014 l'ICAR ha la responsabilità ed il coordinamento dei sistemi informativi archivistici nazionali: il Sistema Informativo degli Archivi di Stato (SIAS); il Sistema Unificato delle Soprintendenze Archivistiche (SIUSA) e il Sistema Guida Generale degli Archivi di Stato (GGASI); i Portali tematici realizzati all'interno del SAN e quelli all'interno degli altri sistemi informativi archivistici nazionali: Antenati - Gli archivi per la ricerca anagrafica; Archivi degli architetti; Archivi d'impresa; Archivi della moda del Novecento; Carte da legare; Rete degli archivi per non dimenticare; Verdi on line.

L'ICAR ha definito e pianificato una strategia di sviluppo di una infrastruttura digitale volta al miglioramento dei servizi per la fruizione del patrimonio archivistico con gli obiettivi di:

- allineamento al Piano Triennale per l'informatica nella PA,

- implementazione di un ecosistema per la fruizione dei beni culturali archivistici,
- definizione di un modello per conservazione del patrimonio informativo digitale.

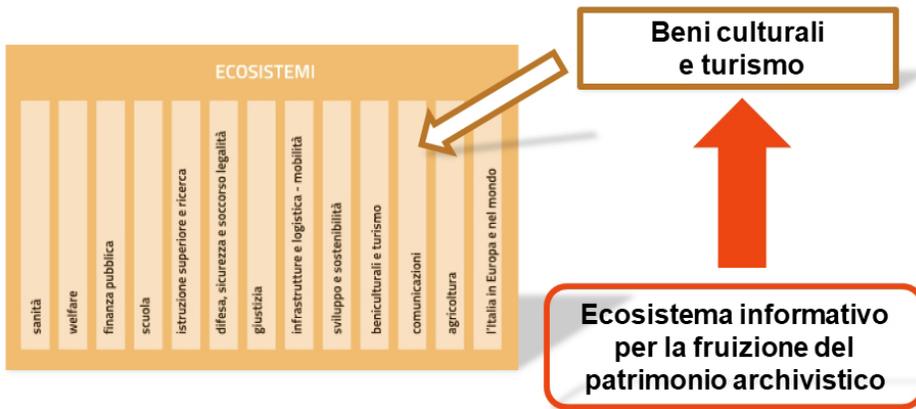


Fig. 1  
Ecosistema informativo per la fruizione del patrimonio archivistico

Coerentemente con questa strategia sono stati individuati:

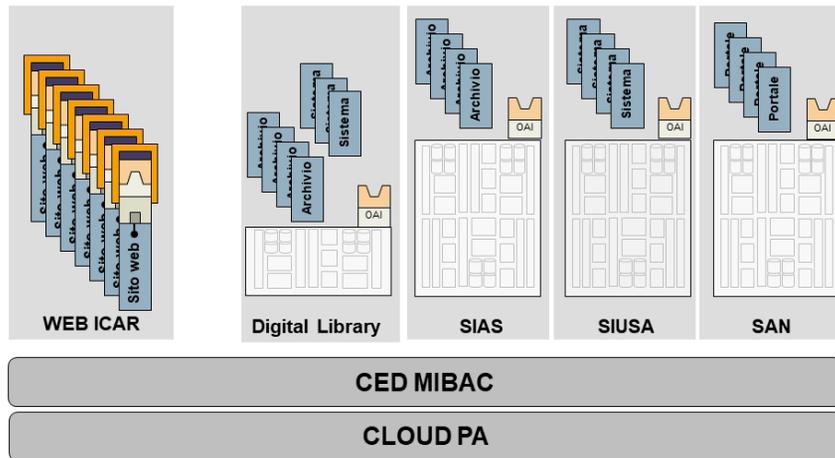
- gli standard di riferimento con architetture di tipo open source e neutralità tecnologica;
- l'uniformità architeturale delle soluzioni di tipo web based con esposizione dei servizi;
- il modello di fruizione specifico per le esigenze di accesso a contenuti archivistici;
- i vocabolari, le ontologie di riferimento e gli authorities file, elementi essenziali per garantire uniformità concettuale degli elementi di raccordo di accesso alle informazioni;
- i web services di accesso alle funzioni essenziali delle varie componenti dell'ecosistema per poter distribuire informazioni qualificate ed autorevoli.

L'ICAR fornirà una infrastruttura nazionale per l'erogazione di servizi per la gestione del patrimonio archivistico con una focalizzazione specifica sulle tematiche della sostenibilità nel tempo della soluzione, sulla fruizione dei contenuti, sulle politiche di accesso alle informazioni archivistiche in possesso degli istituti archivistici e una base per una implementazione di politiche di conservazione del patrimonio informativo digitale.

Tutti gli attuali Sistemi informativi archivistici nazionali saranno evoluti per uniformarsi al modello previsto e quindi integrando servizi di pubblicazione di inventari e digital library specializzate nella gestione del patrimonio digitale. Per supportare i processi di alimentazione dei contenuti e di interoperabilità fra i vari attori di questo ecosistema informativo, sono stati definiti modelli e tracciati di scambio che ne costituiscono l'ossatura concettuale e formale.

I tracciati CAT-SAN e METS-SAN permettono sia di scambiare dati con il Sistema Archivistico Nazionale e sia di pubblicarli con un endpoint Sparql attraverso una ontologia specifica. I nuovi tracciati ICAR li evolvono ampliandone le casistiche d'uso e permettono di gestire le entità che tipicamente caratterizzano i sistemi descrittivi sviluppati nel nostro paese (complessi archivistici, unità/sottounità archivistiche, unità documentarie, soggetti produttori, soggetti conservatori, strumenti di ricerca, entità/voci d'indice).

Fig. 2  
Modello generale  
dei servizi



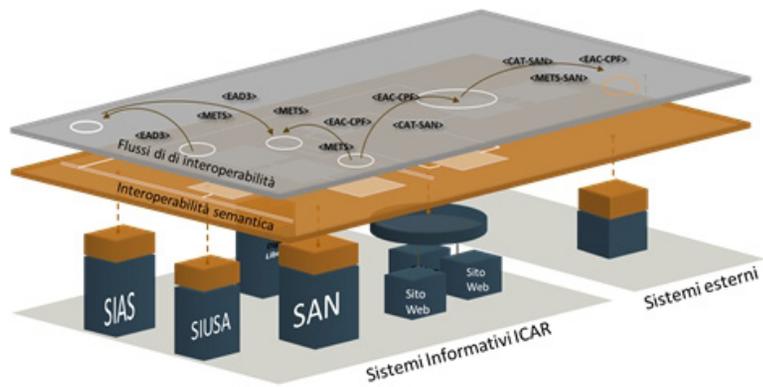
In questo modo permettono la più ampia interoperabilità possibile fra i sistemi di descrizione archivistica. Sono stati allineati alle ultime versioni degli standard: EAD3 per i complessi archivistici e per gli strumenti di ricerca; EAC-CPF per la descrizione dei soggetti produttori; SCONS2 per la descrizione dei soggetti conservatori di archivi.

## 2. Servizi

Il modello di evoluzione si pone come obiettivo la disponibilità di software e servizi gratuiti per la fruizione del patrimonio archivistico e li suddivide nei seguenti servizi e sottoservizi specifici:

- Fornitura di sistemi informativi condivisi per la descrizione archivistica SIAS e SIUSA
  - Disponibilità di Web services per la condivisione dei contenuti
  - Costante allineamento ai tracciati e ai modelli concettuali dell'ICAR
  - Software disponibile gratuitamente e mantenuto dall'ICAR
- Disponibilità di software open source per l'inventariazione
  - Costante allineamento ai tracciati e ai modelli concettuali dell'ICAR
  - Software disponibile gratuitamente e mantenuto dall'ICAR
- Fornitura di una Digital library archivistica condivisa open source
  - Servizio generale fornito per ogni singolo Archivio di Stato
  - Costante allineamento ai tracciati e ai modelli concettuali dell'ICAR
  - Software disponibile gratuitamente e mantenuto dall'ICAR
- Fornitura di un servizio condiviso per la gestione di siti web open source
  - Template predefiniti per il Content Management System Typo 3
  - Soluzione disponibile gratuitamente e mantenuta dall'ICAR
- Portali tematici nazionali
  - Implementazione del Sistema Archivistico Nazionale,
  - Implementazione dei Portali nazionali con contenuti e fonti archivistiche tematizzate
  - Costante allineamento ai tracciati e ai modelli concettuali dell'ICAR

Fig. 3  
Modello generale  
dei flussi fra i  
sistemi



### 3. Elementi chiave del modello

Il modello di evoluzione dei propri sistemi informativi adottato dall'ICAR deve essere sostenibile sia economicamente che organizzativamente con la necessità di un consolidamento nel tempo e con la necessaria capacità di adattarsi a scenari diversi.

I punti chiave del modello sono:

- Consolidamento e ottimizzazione delle infrastrutture informatiche con condivisione delle risorse
  - Utilizzo del CED del MiBAC
  - Utilizzo di fornitori esterni per le esigenze elaborative specifiche e per lo storage
- Software open source distribuito alla comunità di riferimento
  - SIAS, SIUSA, Archimista/Archivista, metaFAD
  - modelli CMS Typo3
- Interoperabilità concettuale degli elementi del modello
  - Continuo allineamento agli standard internazionali
  - Definizione di standard di dominio nazionali e loro rappresentazione con schemi xml e ontologie
  - Gestione e creazione di Authorities files di entità specifiche del dominio archivistico
- Formazione
  - Sostegno alla comunità open source di riferimento per i software individuati
  - Diffusione degli standard e dei modelli concettuali

### 4. Target di riferimento

Il target dei servizi archivistici è l'insieme della rete degli Archivi di Stato e delle Soprintendenze Archivistiche: 120 organizzazioni sparse sul territorio nazionale che forniscono l'accesso e la gestione del patrimonio archivistico nazionale.

I contenuti dell'intero ecosistema saranno valutati in base alla normativa attuale in ottemperanza alle norme in materia di dati personali e alle "Regole deontologiche per il trattamento a fini di archiviazione nel pubblico interesse o per scopi di ricerca storica", ai sensi dell'art. 20, comma 4, del d.lgs. 10 agosto 2018, n. 101 - 19 dicembre 2018).

I servizi saranno riprogettati in funzione di questa analisi e saranno predisposti meccanismi di identificazione degli utenti che saranno preliminari all'accesso alla documentazione digitale. I servizi di identificazioni individuati sono due: la Piattaforma abilitante del Sistema Pubblico di Identità Digitale ed il sistema di identificazione IDEM (Università e Ricerca).

### 5. Infrastrutture abilitanti

Elementi cardine per il corretto funzionamento di ecosistemi informativi come quello presentato sono la disponibilità di infrastrutture abilitanti: la condivisione e la trasmissione di grandi quantità di dati, la velocità di accesso ai sistemi, l'accesso a sistemi di conservazione digitale.

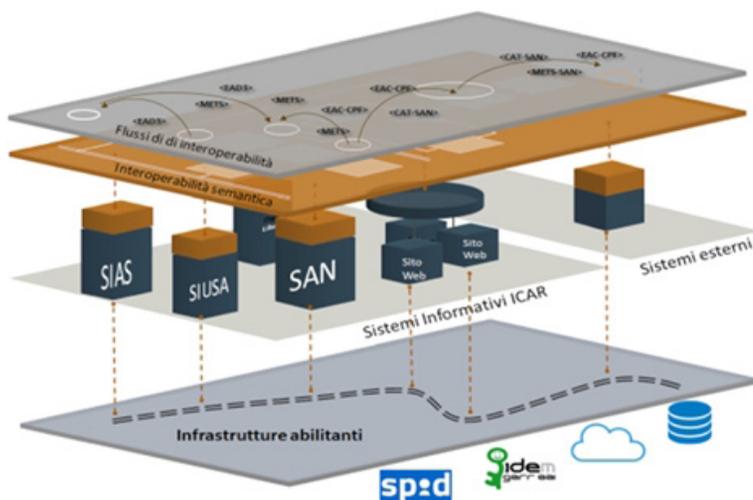


Fig. 4  
Modello generale delle infrastrutture abilitanti

### Autore



Costantino Landino - [costantino.landino@gmail.com](mailto:costantino.landino@gmail.com)

Project manager e consulente IT. Ha sempre lavorato alla progettazione e alla gestione dei sistemi informativi nei settori pubblico e privato. Nel settore dei beni culturali, ha gestito progetti in vari settori e ha seguito lo sviluppo di sistemi di gestione dei documenti, sistemi di archiviazione e biblioteche digitali. Esperto in linguaggi XML, metadati, Linked Open Data e ontologie. Si occupa di conservazione digitale a lungo termine e archiviazione web

# Costruzione di un vocabolario controllato per la terminologia sulla Cybersecurity

Claudia Lanza

Università della Calabria

**Abstract.** Questo articolo si propone di presentare la realizzazione di un primo prototipo di vocabolario di termini controllato, un thesaurus, in lingua italiana per la sicurezza informatica all'interno di un progetto in collaborazione con l'Istituto di Informatica e Telematica (IIT) del Consiglio Nazionale delle Ricerche (CNR), Cyber-Lab – Osservatorio sulla Cybersecurity (OCS). L'obiettivo principale è stato quello di fornire una risorsa in grado di comprendere quanta più informazione autorevole sul mondo della sicurezza informatica nonché offrire una piattaforma di condivisione per far meglio comprendere le interconnessioni terminologiche proprie di questo lessico specialistico sia agli utenti esperti che comuni.

**Keywords.** Cybersecurity; Thesaurus; Gestione della conoscenza; Formazione; Condivisione e riuso dei Dati.

## Introduzione

Quando ci si confronta con domini che presentano un lessico specializzato, spesso risulta difficile comprendere pienamente il modo in cui i suoi concetti sono interconnessi tra di loro. È in questi casi che l'utilizzo di risorse di gestione e rappresentazione della conoscenza (SKOS) può fornire un importante contributo alla comprensione di un lessico tecnico. Tra le risorse semantiche impiegate ai fini di un orientamento più facilitato alla conoscenza di ambiti specialistici, e alla condivisione dei dati, c'è il thesaurus. Come Broughton sottolinea, il thesaurus è spesso utilizzato sia per la rappresentazione dei termini più significativi associati a un dominio di studio che per l'indicizzazione dei documenti ad esso relativi.

La prima parte di questo articolo sarà dedicata alla presentazione della metodologia utilizzata per il recupero dei testi riferiti al mondo della Cybersecurity, di fatto la costruzione del corpus. La seconda parte si concentrerà sulla collaborazione con il gruppo di esperti di dominio, il gruppo dell'IIT presso il CNR di Pisa, con i quali si è avviato un progetto cooperativo per la costruzione di una piattaforma web rivolta a esperti del settore e a utenti comuni ai fini dello sviluppo di una risorsa online sulla sicurezza informatica; seguirà una breve descrizione della strategia di mappatura semantica tra i termini candidati per far parte del thesaurus italiano sulla sicurezza informatica e le tassonomie presenti negli standard di riferimento per la comunità degli esperti in Cybersecurity, i.e., NIST e ISO 27000:2016.

Infine l'articolo si concluderà con la delineazione della costruzione del thesaurus come guida e formazione per l'utente alla comprensione di tecnicismi.

## 1. Costruzione del corpus

Per la costruzione del corpus è stato seguito il sistema della gerarchia delle fonti secondo cui è preferibile organizzare il reperimento dei documenti sul settore scientifico di interesse partendo prima dal livello istituzionale-normativo. Il corpus ha iniziato ad essere popolato di decreti legislativi, norme, direttive ministeriali, regolamenti, tutti riferiti al campo della sicurezza informatica in lingua italiana. Successivamente, il bacino è stato allargato anche alle riviste di settore che hanno contribuito ad espandere sempre di più la terminologia di dominio. Infatti, la Cybersecurity presenta un lessico molto variegato, caratterizzato da un'importante impronta legislativa che ne gestisce la regolamentazione, nonché da tecnicismi adoperati da esperti di settore su riviste di dominio. Due delle principali fonti italiane divulgative analizzate sono le riviste "GNOSIS" e "Cybersecurity Trends", e, tra le linee guida, quelle pubblicate dal Computer Emergency Response Team (CERT), così come glossari specifici, e.g. "Glossario Intelligence" e "Threatsaurus".

## 2. Affiancamento con gli esperti di dominio

I documenti costituenti del corpus di partenza sono stati sottoposti a un processo di trattamento linguistico per estrarre i termini più rappresentativi che sarebbero diventati parte del thesaurus italiano sulla sicurezza informatica. Il software scelto per l'estrazione terminologica è Text To Knowledge (T2K) che permette di ottenere in output una lista di termini significativi in base alla formula Term Frequency–Inverse Document Frequency (TF-IDF): vengono ordinati in modo decrescente i termini considerati più significativi all'interno dei documenti dati in input come corpus di partenza.

A partire da questo elenco iniziale di termini candidati si è avviata una prima forma di collaborazione con gli esperti di dominio. Come Schultz evidenzia, per il lavoro del terminologo diventa molto importante la fase di affiancamento con il gruppo di esperti di dominio in quanto sono proprio questi ultimi a garantire un livello di valutazione appropriata circa la base contenutistica ottenuta dai software di estrazione semantica. I termini sono la rappresentazione linguistica dei concetti propri di un dominio specifico, e il thesaurus ha l'obiettivo di fornire una sistematica rappresentazione della struttura di conoscenza propria a determinati domini specialistici. Pertanto, la figura degli esperti diviene fondamentale ai fini della selezione più accurata dei termini che rappresentano il dominio. Con il gruppo di esperti sono stati definiti sia la determinazione di rapporti semantici esistenti tra i termini estratti dai documenti del corpus che l'inserimento di nuovi contenuti ai fini dell'arricchimento tesaurale.

### 2.1 Comparazione con gli standard di riferimento

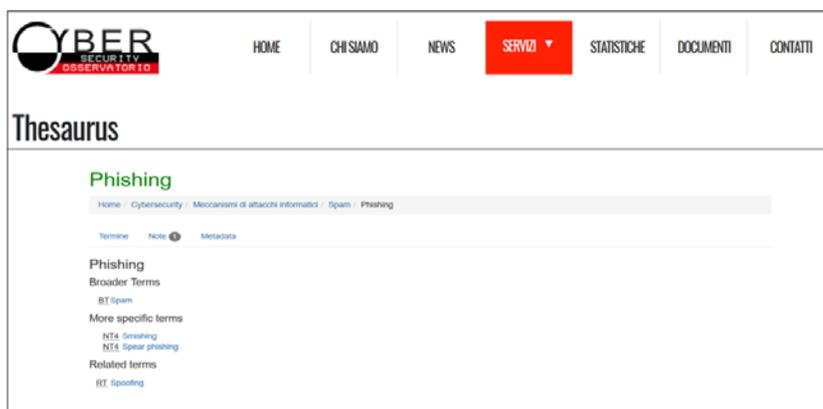
Per ottenere un alto grado di accuratezza e precisione terminologica nella struttura tesaurale italiana per la sicurezza informatica, è stato avviato a monte un processo di mappatura tra le tassonomie presenti negli standard ufficiali di riferimento per gli esperti di Cybersecurity e le liste ottenute dall'estrazione terminologica con T2K. Le due banche dati ufficiali con cui sono stati mappati i termini derivati dal corpus di partenza sono il vocabolario e il glossario contenuti rispettivamente negli standard ISO/IEC 27000:2016

e NIST 7298. È stata verificata la presenza dei termini interni ai primi due elenchi sulla lista dei termini ottenuti con T2K per convalidare la copertura semantica data dai termini rappresentativi candidati e filtrati insieme agli esperti.

### 3. Conclusioni

L'obiettivo principale alla base dello sviluppo del thesaurus italiano per la Cybersecurity è stato quello di fornire uno tipo di supporto linguistico non ancora, secondo ricerche effettuate e sulla base delle nostre conoscenze, esistente ufficialmente nell'ambito italiano. Il thesaurus è strutturato sulla base di relazioni semantiche di base che gestiscono i rapporti tra i termini: la relazione di gerarchia che regola i rapporti tra i termini più generici (BT) e quelli più specifici (NT); la relazione di sinonimia tra i termini preferiti (USE) e quelli non preferiti (UF) e la relazione di associazione (RT). Questa struttura di gestione della conoscenza permette una visualizzazione dinamica dei termini costitutivi di un dominio tecnico, come quello della Cybersecurity. A fini esemplificativi si veda la Fig. 1 in cui il termine "Phishing" viene rappresentato in base a una serie di rapporti semantici che intrattiene con altri termini interni al thesaurus: "Phishing" è un termine più specifico (NT) di "Spam" (BT), che, a sua volta, è un termine più specifico di "Meccanismi di attacchi informatici".

Fig. 1  
"Phishing",  
Thesaurus italiano  
per la Cybersecurity



### Riferimenti bibliografici

A. Milers, S. Bechoofer, "SKOS Simple Knowledge Organization System Reference", W3C Recommendation 18 August 2009.

C. Sammut and G. I. Webb, "Encyclopedia of Machine Learning" (1st ed.). Springer Publishing Company, Incorporated, 2011.

CERT Nazionale Italia, < <https://www.certnazionale.it/category/linee-guida/> > (ultimo accesso 10/04/2019).

Claire K. Shultz, Wallace L. Schultz, and Richard H. Orr, "Evaluation of Indexing by Group Consensus" (Final Report, Contract No OEC 1-7-070622-3890), Bureau of Research Office of Education, U.S. Department of Health, Education and Welfare, August 30, 1968, p. 40.

Cybersecurity Osservatorio, <<https://www.cybersecurityosservatorio.it/>> (ultimo accesso 10/04/2019).

Cybersecurity Trends, < <https://www.cybertrends.it/>> (ultimo accesso 10/04/2019).

F. Dell'Orletta, G. V. (2014). "T2K<sup>2</sup>: a System for Automatically Extracting and Organizing Knowledge from Texts", in Proceedings of 9th Edition of International Conference on Language Resources and Evaluation (LREC 2014).

G. Zagrebelsky, "Il sistema costituzionale delle fonti del diritto", Torino, EGES; UTET, 1984, p. 67.

GNOSIS Rivista italiana di intelligence, <<http://gnosis.aisi.gov.it/gnosis/Start.nsf/pages/home>> (ultimo accesso 10/04/2019).

International Standard ISO/IEC 27000:2016 (E) Information technology – Security techniques – Information security management systems – Overview and vocabulary, Fourth edition 2016-02-05.

J. C. Sager, "A practical course in terminology processing", John Benjamins Publishing Company, 1990.

Presidenza del Consiglio dei Ministri – Sistema di informazione per la sicurezza della Repubblica, "Il linguaggio degli Organismi Informativi", Glossario Intelligence, <<https://www.sicurezzanazionale.gov.it/sisr.nsf/quaderni-di-intelligence/glossario-intelligence.html>> (ultimo accesso Ottobre 2018).

R. Kissler, NISTIR 7298 National Institute of Standards and Technology Interagency or Internal Report 7298r2, 2013, May.

Sophos "Threatsaurus The A-Z of Computer and data security threats", <<https://www.sophos.com/en-us/medialibrary/PDFs/other/sophosthreatsaurusaz.pdf?la=en>> (ultimo accesso Settembre 2018).

W. Broughton, "Costruire Thesauri: strumenti per indicizzazione e metadati semantici" (a cura di) P. Cavaleri, (traduzione di) L. Ballestra e L. Venuti, Milano, EditriceBibliografica, 2008.

## Autrice



**Claudia Lanza** - [c.lanza@dimes.unical.it](mailto:c.lanza@dimes.unical.it)

Dottoranda in ICT presso l'Università della Calabria, Dipartimento DIMES, dal 2017. Ha compiuto gli studi specialistici a Milano presso l'Università Cattolica del Sacro Cuore in "Scienze del linguaggio, terminologie e tipologie dei testi" dove si è laureata nel 2016.

Attualmente è PhD Visiting fino al 2020 presso l'Université de Nantes, Laboratoire LS2N, all'interno del gruppo TALN (Traitement automatique du langage naturel). Il suo campo di ricerca comprende lo sviluppo di risorse per l'organizzazione e rappresentazione di lessici specialistici. In particolar modo, le aree di investigazione riguardano la costruzione di risorse semantiche come thesauri e ontologie e lo studio di metodi di linguistica computazionale per l'automatizzazione di operazioni semantiche.

# Una soluzione basata sulla tecnologia blockchain per la validazione di attestati e titoli rilasciati tramite piattaforma di e-learning

Nicola Cannistrà, Fabio Cordaro, Fabrizio La Rosa, Francesco La Rosa, Roberta Maisano, Umberto Ruggeri, Salvatore Todaro, Riccardo Uccello

Università degli Studi di Messina

**Abstract.** Negli ultimi anni, le criptovalute e la tecnologia blockchain a supporto delle transazioni hanno attirato un crescente interesse. Spesso, le criptovalute come bitcoin vengono associate erroneamente alle blockchain come un unico concetto. In realtà la blockchain rappresenta uno strumento che prevede molteplici applicazioni e il cui scopo è quello di creare asset digitali unici dotati di particolari caratteristiche di integrità, affidabilità e non ripudio. Questa tecnologia cambierà radicalmente la modalità di gestione delle attività amministrative, finanziarie e di management con una gestione trasparente e democratica. In questo articolo proponiamo una soluzione basata sulla blockchain, ed in particolare sugli smart contract, finalizzata alla verifica dei certificati rilasciati come attestazione del successo formativo raggiunto o della partecipazione a corsi tenuti in modalità e-learning.

**Keywords.** Blockchain, ledger, smart contract, Ethereum

## Introduzione

La necessità di avere un sistema in grado di garantire l'immutabilità delle informazioni registrate e la sicurezza delle transazioni, con il massimo livello di condivisione e trasparenza possibile, ha spinto gli addetti ai lavori verso un interesse sempre maggiore nei confronti della blockchain (Zheng et al. 2017). La blockchain permette la creazione di un database distribuito in cui le transazioni ed i dati vengono condivisi tra più nodi presenti in una rete (figura 1). Si tratta di un sistema strutturato a blocchi, concatenati l'uno all'altro per formare una lunga catena contenente tutte le transazioni e condivisa da tutti i nodi che sono deputati alla validazione, controllo ed approvazione.

L'elenco dei dati rappresentati dai blocchi della catena è in continua crescita e, grazie a un algoritmo di consenso distribuito su una rete peer-to-peer decentralizzata, le transazioni vengono verificate continuamente prima di aggiungere blocchi al libro mastro pubblico (Distributed Ledger).

Il processo di verifica consente che le transazioni appena aggiunte non siano in contrasto con le transazioni confermate nella blockchain e mantengono l'ordine cronologico corretto. La validazione avviene tramite un processo crittografico delle operazioni racchiuse nella blockchain. Ogni operazione avrà una sorta di impronta registrata nel blocco con un timestamp che risulta quindi immodificabile.

Possiamo affermare che con l'utilizzo della firma digitale, della crittografia e della

Fig. 1  
Una rappresentazione  
dei nodi distribuiti  
a supporto di una  
blockchain



blockchain (Wright et al. 2015) il processo di validazione presenta le caratteristiche di integrità, trasparenza e non ripudio, necessarie per l'utilizzo in procedimenti affidabili e legalmente validi. Gli studi sulle potenzialità delle blockchain (De La Rosa et al. 2017) sono tuttora in corso e molto vivo è anche il dibattito su quale tipo di blockchain configurare, tuttavia esistono già diversi casi di possibile utilizzo (Faioli et al. 2016; Ølnes et al. 2017). Gli smart contract (Alharby et al. 2017) sono contratti scritti in un linguaggio di programmazione in modo da determinare, automaticamente, l'esecuzione delle clausole contrattuali all'avverarsi delle determinate condizioni inserite nel contratto stesso. A differenza di un contratto tradizionale in cui, dopo aver raggiunto un accordo, le parti devono "eseguire" il contratto per renderlo operativo, uno smart contract è autoeseguibile, cioè, una volta che le istruzioni sono state scritte su una blockchain, la transazione avverrà automaticamente quando vengono rilevate delle condizioni appropriate.

In questo articolo proponiamo una soluzione basata su blockchain, ed in particolare sugli smart contract, finalizzata alla verifica dei certificati rilasciati come risultato del successo formativo raggiunto o della partecipazione a corsi erogati tramite una piattaforma di e-learning.

## 1. Il sistema

In questo lavoro i certificati rilasciati ai partecipanti di alcuni corsi di formazione ospitati sulla piattaforma di e-learning dell'Università degli studi di Messina (che ospita corsi universitari, corsi di formazione destinati al personale tecnico amministrativo, master universitari, ecc.) come riconoscimento dei risultati formativi raggiunti vengono archiviati sulla blockchain di Ethereum e, se necessario, se ne può controllare la loro autenticità. La piattaforma Moodle è adottata dall'Università degli Studi di Messina dal 2005. Ai partecipanti al termine dei corsi, o al superamento delle prove d'esame, viene rilasciato un certificato. Gli strumenti di certificazione per Moodle già esistenti (Accredible, Simple Certificate, ecc.) non sono tutti gratuiti e molti non sono compatibili con la versione attuale (ver. 3.7.1). Inoltre, solo uno di questi strumenti (Accredible) ha tratto vantaggio dalla

funzione chiave pubblica-privata della tecnologia blockchain (Mikroyannidis et al. 2018). Ai fini di questo lavoro, una versione custom del modulo di certificazione del Moodle Learning Management System è stato realizzato per funzionare in conformità con uno smart contract definito su blockchain Ethereum. Nonostante ci siano diverse blockchain che supportano gli smart contract, i motivi per preferire l'Ethereum Blockchain sono l'utilizzo da parte di una grande comunità, la possibilità di un rapido accesso a soluzioni di possibili errori, la presenza di un ampio supporto API e la libreria di web3.js o web3.php (Ethereum per PHP) compatibile con Moodle.

### **1.1 Ethereum Blockchain**

Ethereum, lanciata nel 2013 come idea di Vitalik Buterin, diventa operativa nel 2015. La caratteristica che distingue Ethereum dalle altre blockchain è che consente lo sviluppo e l'esecuzione di "smart contract" e "applicazioni autonome distribuite - DApp". Il linguaggio di programmazione utilizzato è chiamato "Solidity". Il programma è compilato e convertito in bytecode e inviato alla blockchain di Ethereum come smart contract. Contratti e applicazioni sulla blockchain vengono eseguiti su una Ethereum Virtual Machine (EVM). L'utilizzo di smart contract sulla blockchain e l'approvazione della transazione comportano costi in base alla dimensione del contratto in bytecode, la quantità di dati inviati e il numero di transazioni. Questi costi sono indicati come Ether o Gas, durante la spedizione e il funzionamento dello smart contract. Una verifica preventiva del corretto funzionamento del sistema avviene in ambienti di test denominati Testnet.

### **1.2 Ambiente di test**

Per sviluppare uno smart contract sulla blockchain di Ethereum è stata installata e configurata l'applicazione Ganache. L'ambiente scelto per lavorare in combinazione con Ganache e per scrivere smart contract è Remix, che è un IDE dotato di interfaccia web. Dopo aver sviluppato lo smart contract di base, il processo di sviluppo del contratto è stato completato utilizzando MyEtherWallet. Lo smart contract, sviluppato con Remix e quindi inviato alla rete di test tramite l'interfaccia di MyEtherWallet, è costituito da due funzioni principali. La prima funzione registra il nome, il cognome, le informazioni sull'identità del partecipante, l'istituto che rilascia il certificato, il codice di autenticazione del documento e il valore hash md5 del documento che è ottenuto da Moodle, sulla blockchain di Ethereum. L'altra funzione consente, fornendo il codice di autenticazione del documento (specificato nel certificato), di verificarne l'autenticità e ottenere informazioni correlate richiamando l'hash della transazione inserita nella blockchain.

### **1.3 Learning System**

Il modulo di certificazione sviluppato consente agli studenti di scaricare i propri certificati direttamente dalla piattaforma Moodle. Le informazioni relative al partecipante vengono elaborate dalla piattaforma ed utilizzate per produrre un file in formato PDF protetto da password. Quando gli studenti scaricano il loro certificato, le informazioni su candidato, istituto e file vengono registrate nella blockchain Ethereum mediante l'esecuzione dello

smart contract. Una volta completata la registrazione, il certificato diventa disponibile per il download e il codice di controllo del documento e il valore hash della transazione del processo vengono scritti in dei campi appositi del profilo dello studente in modo da consentire una verifica successiva.

## 2. Conclusioni

La tecnologia blockchain, mediante l'uso della crittografia, permette la riconciliazione e la replica dei dati, il controllo degli accessi, la trasparenza e la privacy. La blockchain, inoltre, trasforma l'organizzazione e le modalità di prestare il lavoro, incidendo sulla cultura giuridica, determinando una riorganizzazione della pubblica amministrazione e dei relativi modi di erogare servizi. Questo lavoro, grazie alla versatilità della tecnologia blockchain, rivoluziona il modo di tracciare le competenze, proponendo una soluzione innovativa basata sugli smart contract, finalizzata alla verifica dei certificati rilasciati tramite una piattaforma di e-learning.

## Riferimenti bibliografici

Alharby M., van Moorsel A. (2017), Blockchain based smart contracts: A systematic mapping study, fourth International Conference on Computer Science and Information Technology.

De La Rosa J.L., Torres-Padrosa V., El-Fakdi A., Gibovic D., Hornyak O., Maicher L., Miralles F. (2017), A survey of blockchain technologies for open innovation, fourth Annual World Open Innovation Conference.

Faioli M., Petrilli E., Faioli D. (2016), Blockchain, contratti e lavoro. la ri-rivoluzione del digitale nel mondo produttivo e nella pa, Economia and lavoro, Rivista di politica sindacale, sociologia e relazioni industriali, 2, pp.139-158.

Mikroyannidis A., Domingue J., Bachler M., Quick K. (2018), Smart blockchain badges for data science education, IEEE Frontiers in Education Conference.

Ølnes S., Ubacht J., Janssen M. (2017), Blockchain in government: Benefits and implications of distributed ledger technology for information sharing, Government Information Quarterly, 34(3), pp.355-364.

Wright A., De Filippi P. (2015), Decentralized blockchain technology and the rise of lex cryptographia, SSRN Electronic Journal.

Zheng Z., Xie S., Dai H., Chen X., Wang H. (2017), An overview of blockchain technology: Architecture, consensus, and future trends. IEEE International Congress on Big Data, pp. 557-564.

## Autori



Nicola Cannistrà - [nicola.cannistra@unime.it](mailto:nicola.cannistra@unime.it)

APM-GARR per l'Università di Messina. Responsabile dell'U.Op. "Infrastrutture ICT" dell'Università degli Studi di Messina, ha sviluppato competenze in ambito sistemistico, progettazione e gestione reti in ambito MAN, sistemi di WiFi centralizzato, sistemi VoIP, sicurezza e sistemi di autenticazione.

**Fabio Cordaro** - [fabio.cordaro@unime.it](mailto:fabio.cordaro@unime.it)

Responsabile dell'Unità Operativa "Sicurezza e Datacenter" presso l'Università degli Studi di Messina, Technical Contact per il dominio della stessa università, ha sviluppato competenze nell'amministrazione di sistemi Unix-like, progettazione e gestione reti, amministrazione di servizi di rete, implementazione di applicativi web.



**Fabrizio La Rosa** - [fabrizio.larosa@unime.it](mailto:fabrizio.larosa@unime.it)

Laureato in Matematica c/o l'Università degli Studi di Messina. Negli ultimi anni ha ricoperto il ruolo di responsabile delle infrastrutture ICT, reti e datacenter per l'Università degli Studi di Messina ed il Comune di Milano e quello di APM per l'Ateneo messinese.



**Francesco La Rosa** - [francesco.larosa@unime.it](mailto:francesco.larosa@unime.it)

Ha conseguito un dottorato di ricerca in Computer Science c/o l'Università degli Studi di Messina. E' coautore di decine di articoli pubblicati su atti di convegno e riviste internazionali. E' responsabile dell'Unità di Staff "Servizi di Rete" presso l'Università degli Studi di Messina.



**Roberta Maisano** - [roberta.maisano@unime.it](mailto:roberta.maisano@unime.it)

Laureata in Engineering and Computer Science presso l'Università degli Studi di Messina dove ha conseguito anche un Dottorato di Ricerca. E' coautrice di diversi articoli scientifici su tematiche di data science, image processing e machine learning. E' vice-responsabile dell'Unità di Staff Servizi di Rete presso l'Università degli Studi di Messina.



**Umberto Ruggeri** - [umberto.ruggeri@unime.it](mailto:umberto.ruggeri@unime.it)

Laureato in Ingegneria Elettronica presso l'Università degli Studi di Messina. Responsabile dell'Unità Organizzativa Sicurezza, Servizi Software e Gestione Energetica presso l'Università degli Studi di Messina. Ha sviluppato competenze in ambito sistemistico, virtualizzazione di sistemi, progettazione di reti, sicurezza e sistemi di autenticazione.



**Salvatore Todaro** - [salvatore.todaro@unime.it](mailto:salvatore.todaro@unime.it)

Vice responsabile dell'Unità Operativa "Sicurezza e Datacenter" presso l'Università degli Studi di Messina, si occupa di gestione delle identità, di sicurezza, gestione sistemi e infrastrutture di virtualizzazione. Ricopre il ruolo di referente tecnico di Ateneo e di membro del Comitato Tecnico Scientifico per la Federazione di Identità IDEM GARR AAI.



**Riccardo Uccello** - [riccardo.uccello@unime.it](mailto:riccardo.uccello@unime.it)

Laureato in Fisica presso l'Università degli Studi di Messina. Ha ricoperto negli ultimi anni ruoli di responsabilità nel settore dell'ICT, prima presso l'Università Mediterranea ultima-mente presso il Centro Informatico dell'Università degli Studi di Messina. Ricopre il ruolo di APA-GARR per l'Ateneo messinese.



# Quanti petabyte occorrono per vedere un buco nero?

Elisabetta Liuzzo<sup>1,2</sup>, Kazi Rygl<sup>1,2</sup>

<sup>1</sup>INAF- Istituto di Radioastronomia, <sup>2</sup>Italian ALMA Regional Centre

**Abstract.** Il 10 aprile scorso il Consorzio dell' Event Horizon Telescope (EHT) ha pubblicato la prima immagine dell'ombra del buco nero supermassiccio al centro della galassia M87. In questo lavoro, mostriamo come l' EHT sia riuscito in questa impresa. Discutiamo inoltre il percorso dei dati scientifici di volume di ~ 4 PB dagli osservatori fino all'immagine finale di soli qualche MB. Riportiamo poi i principali risultati astrofisici di tale scoperta e finiamo con illustrare i lavori in corso e futuri.

**Keywords.** buco nero, orizzonte degli eventi, galassia attiva, M87, Event Horizon Telescope

## Introduzione

I buchi neri sono gli oggetti astrofisici più estremi previsti dalla Relatività Generale (RG, Einstein 1915). Si tratta di corpi celesti con un campo gravitazionale molto intenso caratterizzati da una superficie, chiamata l'orizzonte degli eventi, oltre la quale, né la materia né i fotoni possono sfuggire (e quindi arrivare a noi), il che limita l'informazione che possiamo avere di tali regioni centrali. Al di fuori dell'orizzonte degli eventi, i fotoni vengono attratti e deviati per l'intenso campo gravitazionale che deforma fortemente lo spazio-tempo, tramite il fenomeno di lensing gravitazionale (Luminet 1979). Secondo la RG, l'emissione di tali fotoni ha forma di ciambella con al centro un buco, cioè assenza di segnale (l'ombra) attorno all'orizzonte degli eventi.

Storicamente, le osservazioni dei buchi neri sono state indirette, ovvero tramite lo studio degli effetti che questi provocano sugli oggetti che vi ruotano attorno (Genzel et al. 2010). Per lungo tempo, l' anello di fotoni che circonda l'ombra del buco è stato impossibile da rilevare. Infatti esso è di emissione così debole e su scale angolari così piccole da essere osservabile solo tramite strumenti che lavorano in banda millimetrica e che usano tecniche di interferometria a lunga base (Falcke et al. 2000). Negli ultimi 20 anni, diversi sforzi sono stati fatti per costruire uno strumento del genere. Questi sono culminati nella creazione dell' Event Horizon Telescope (EHT) che nel 2017 è stato così potente e sensibile da riuscire ad osservare due buchi neri supermassicci con le dimensioni dell'orizzonte degli eventi più grandi: quello della nostra Galassia Sagittarius (Sgr) A\* e di M87, galassia a 55 milioni di anni luce da noi.

## 1. EHT: un telescopio grande come la Terra

L'astrofisica studia l'Universo attraverso i telescopi il cui potere risolutivo dipende proporzionalmente dal diametro dell'antenna ed inversamente dalla lunghezza d'onda a cui osserva. EHT opera nella banda millimetrica ed è stato ideato da una collaborazione internazionale

di più di 200 membri con l'obiettivo di produrre immagini dei buchi neri su scale dell'orizzonte degli eventi. Si tratta di un telescopio 10 volte più sensibile dei suoi predecessori e con una risoluzione incredibile, equivalente a quella che occorre per distinguere una pallina da tennis sulla Luna.

Per raggiungere questi dettagli estremi, EHT utilizza la tecnica dell'interferometria a lunga base (VLBI), cioè antenne dislocate nel mondo che sintetizzano un telescopio grande come la Terra. La prima immagine di un buco nero pubblicata lo scorso aprile deriva dalla campagna EHT del 2017 condotta da 8 telescopi in 6 siti geografici diversi (Fig.1) che hanno osservato a 1.3 mm con una larghezza di banda di 4 GHz e un rate di registrazione dei dati di 32 Gbps tramite tecnologia Mark 6 VLBI (Whitney et al. 2013). Le attuali capacità di EHT sono ora 64 Gbps, 10 volte maggiori dei tradizionali interferometri a lunga base, a testimonianza che EHT è uno strumento in continua evoluzione e miglioramento. Il più sensibile componente di EHT è l'Atacama Large Millimetre/submillimetre Array (ALMA) situato in Chile. Senza il suo contributo, non sarebbe stato possibile raggiungere il dettaglio e la sensibilità per poter risolvere l'anello (Fig. 2).



Fig. 1  
Distribuzione geografica delle antenne EHT che hanno partecipato alla campagna osservativa del 2017.

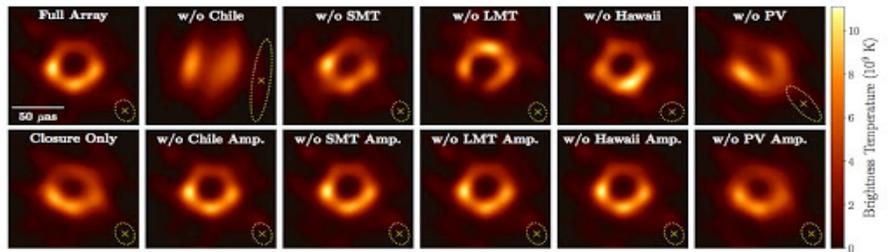
## 1.1. Quanti dati?

Le osservazioni EHT richiedono la gestione di una notevole quantità di dati. Ciascun antenna osserva simultaneamente lo stesso oggetto celeste ed i dati vengono registrati su hard disk. Durante la campagna del 2017, le 8 antenne hanno prodotto circa 4 PB di dati in totale per le 5 giorni osservativi.

Trattandosi di PB, i dati sugli hard disk vengono fisicamente spediti (e non trasferiti tramite internet) ai due correlatori (al Max Planck Institute for Radio Astronomy e al MIT Haystack Observatory) di tipo DIFX (Deller et al. 2011) dove vengono tra loro combinati (allineati in tempo) e mediati in tempo e frequenza raggiungendo dimensioni dell'ordine di 10-100 GB per sorgente per giorno.

Fig. 2

Da sinistra a destra, immagini del buco nero di M87 ottenute usando tutte le antenne EHT e via via sottraendo un'antenna (EHT Collaboration 2019d). Senza ALMA (pannello w/o Chile) l'anello non è distinguibile.



Terminata la correlazione, i dati vengono passati ai gruppi di lavoro che si occupano della loro calibrazione. L'emissione della sorgente registrata dai singoli telescopi non è tale e quale quella emessa dalla sorgente, ma risulta distorta per gli effetti dell'atmosfera (che varia da sito a sito) e per gli errori strumentali. Queste distorsioni devono essere corrette proprio in questa fase. Qui un'ulteriore media dei dati è effettuata ottenendo pacchetti di ~1 GB per sorgente per giorno.

L'ultima fase, detta di imaging, è quella che dai dati ripuliti produce l'immagine finale (~ MB per sorgente) attraverso operazioni matematiche che implicano anche trasformate di Fourier inverse.

## 1.2 Come è stata prodotta l'immagine?

EHT è uno strumento rivoluzionario che osserva con una sensibilità e risoluzione angolare senza precedenti. Per le sue caratteristiche (per es. antenne di diametro diverso poste in siti con climi differenti, PB di dati prodotti) ha necessitato lo sviluppo di software innovativi sia per la calibrazione che imaging (EHT collaboration 2019c, d). Questo ha richiesto del tempo: infatti tra l'osservazione (aprile 2017) e la pubblicazione (aprile 2019) della prima immagine di un buco nero ci sono voluti 2 anni.

Per assicurare l'affidabilità dei risultati ottenuti, è stato scelto anche un approccio metodologico particolare. Sono stati formati 3 gruppi di lavoro indipendenti (~10 persone ciascuno) per la fase di calibrazione e 3 per la parte di imaging che hanno lavorato separatamente senza interazione, per evitare possibili influenze umane sul processamento dei dati. Ciascun gruppo di lavoro ha sviluppato i propri strumenti di analisi: indipendentemente dai metodi usati, i risultati sono tra loro in accordo, come mostra Fig.3 per l' imaging.

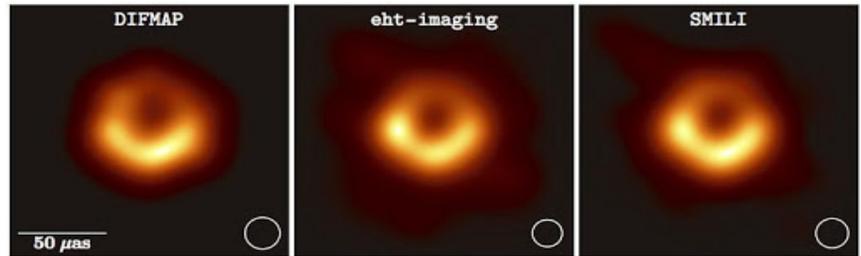
Il nostro contributo è stato quello di contribuire allo sviluppo della pipeline di calibrazione rPICARD (Janssen et al. 2019), scritta in python che utilizza CASA (McMullin et al. 2007), il software più recente ed innovativo usato per la riduzione dei dati radioastronomici, in cui sono state recentemente implementate le funzioni (tra cui l'algoritmo fringefit, van Bemmelen et al. 2019 in prep.) per il processamento dei dati che usano la tecnica interferometrica a lunga base, quali quelli EHT.

## 2. La prima immagine di un buco nero: la galassia M87

La prima immagine pubblicata dal Consorzio EHT lo scorso aprile, precisamente 2 anni

Fig. 3

Immagine del buco nero di M87 prodotta dai gruppi di lavoro che si sono occupati dell'imaging (EHT collaboration 2019d). La sbarra a sinistra indica la scala angolare e il cerchio in basso a destra rappresenta la risoluzione angolare dell'osservazione.



dopo le osservazioni, è stata quella del buco nero supermassiccio al centro della galassia M87. Si tratta dell'immagine in banda millimetrica a più alta risoluzione angolare mai realizzata (20 micro-arcosecondi). In Fig.3, in arancione sono rappresentate le zone che emettono, in nero quelle in cui vi è assenza di segnale. L'emissione è chiaramente a forma di anello. Si tratta di plasma molto caldo ( $T > 10^9 \text{K}$ ) e magnetizzato che ruota attorno al buco nero centrale nel disco di accrescimento (alcune simulazioni sono in accordo con contributo emissivo anche da parte dei getti). Questo plasma incandescente emette per meccanismo di sincrotrone e possiede al suo interno zone di maggiore brillantezza rispetto alle altre (in giallo), indicando la presenza di particelle che si muovono a velocità prossime a quelle della luce. La forma di questo anello è dovuta alla curvatura dello spazio-tempo, il lensing gravitazionale, per la presenza della massa enorme del buco nero centrale, e circonda la famosa ombra del buco nero.

L'immagine costituisce la prima prova diretta dell'esistenza dei buchi neri ed ha permesso di testare la validità della RG rispetto alle teorie alternative anche in ambienti dai campi gravitazionali estremi.

La dimensione dell'anello ha permesso di stimare la massa del buco nero di M87 con grande precisione:  $6.5(\pm 0.7) \times 10^9$  masse Solari (EHT Collaboration 2019f). L'anello non è poi perfettamente simmetrico: la sua ellitticità rivela che il buco nero di M87 appartiene alla famiglia dei buchi neri rotanti detti di Kerr (EHT Collaboration 2019e). Infine, la forma ad anello dell'emissione esclude che si tratti di un buco nero nudo o un wormhole (EHT Collaboration 2019a).

### 3. Conclusioni

Se il 10 aprile scorso il Consorzio EHT ha pubblicato la prima immagine di un buco nero, quello della galassia M87, ora il lavoro in corso ha l'obiettivo di produrre il primo video di un buco nero, quello della nostra Galassia, Sgr A\*. Questo è stato osservato insieme a M87 nel 2017, ma, poiché estremamente variabile e in una zona del cielo con mezzo interstellare particolarmente denso, i dati richiedono un'analisi innovativa che dovrebbe concludersi entro qualche mese.

EHT è uno strumento in continua evoluzione che vuole spingersi ad osservare sempre con più elevato dettaglio. Nuovi telescopi sono stati (e verranno) inseriti nell'array e ricevitori a più alta frequenza (345 GHz) sono in via di sviluppo. Infine, lo scorso agosto, il Con-

sorzio EHT ha sottoscritto un ESA Voyage 2050 White Paper (Gurvits et al. 2019) per la costruzione di un'antenna radio da inviare nello Spazio, il che costituirebbe l'ultima frontiera in termini di risoluzione angolare dell'interferometria a lunga base nel millimetrico.

### Riferimenti bibliografici

Deller A.T., Brisken W.F., Philips C.J., et al 2011, DiFX-2: A More Flexible, Efficient, Robust, and Powerful Software Correlator, *PASP*, 123, 275

EHT Collaboration, 2019a, First M87 Event Horizon Telescope results. I. The shadow of the Supermassive Black Hole, *ApJ* 875, L1

EHT Collaboration, 2019b, First M87 Event Horizon Telescope results. II. Array and Instrumentation, *ApJ* 875, L2

EHT Collaboration, 2019c, First M87 Event Horizon Telescope results. III. Data Processing and Calibration, *ApJ* 875, L3

EHT Collaboration, 2019d, First M87 Event Horizon Telescope results. IV. Imaging the Central Supermassive Black Hole, *ApJ* 875, L4

EHT Collaboration, 2019e, First M87 Event Horizon Telescope results. V. Physical Origin of the Asymmetric Ring., *ApJ* 875, L5

EHT Collaboration, 2019f, First M87 Event Horizon Telescope results. VI. The Shadow and Mass of the Central Black Hole, *ApJ* 875, L6

Einstein A., 1915, *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften*, Berlin

Falcke, H. Melia, F., & Agol, E. 2000, Viewing the Shadow of the Black Hole at the Galactic Center, *ApJL*, 528, L13

Janssen M., Goddi C., van Bemmelen I., et al., 2019, rPICARD: A CASA-based calibration pipeline for VLBI data, *A&A* 626, A75

Genzel R., Eisenhauer F., Gillessen S., 2010, The Galactic Center massive black hole and nuclear star cluster, *Reviews of Modern Physics*, vol. 82, Issue 4, pp. 3121-3195

Gurvits L.I., Paragi Z., Casasola V., et al 2019, TeraHertz Exploration and Zooming-in for Astrophysics (THEZA): ESA Voyage 2050 White Paper, arXiv:1908.10767

Luminet, J.-P, 1979, Image of a spherical black hole with thin accretion disk, *A&A*, 75, 228

McMullin J.P., Waters B., Schiebel D., et al., 2007, *CASA Architecture and Applications*, ASP Conference Series Vol. 376, Tucson, Arizona.

Whitney A.R., Beaudoin C.J., Cappallo R., et al. 2013, Demonstration of a 16 Gbps Station-1 Broadband-RF VLBI System, *PASP*, 125, 196

### Contenuto del text box:

Anche l'Italia ha delle grandi antenne radio: due con diametro di 32m (a Medicina e Noto) ed una di 64m (in Sardegna). Queste antenne vengono anche utilizzate con la stessa tecnica di interferometria a lunga base nel European VLBI Network ([www.evlbi.org](http://www.evlbi.org)). Le

condizioni climatiche non permettono alle antenne italiane di osservare alle lunghezze d'onda del EHT (1mm).

## Autrici



Elisabetta Liuzzo - [elisabetta.liuzzo@inaf.it](mailto:elisabetta.liuzzo@inaf.it)

Lavora come ricercatrice all'INAF-IRA a Bologna. E' membro dell' EHT Consortium ed ha contribuito alla pubblicazione nell'aprile scorso della prima immagine del buco nero di M87, partecipando al gruppo di lavoro per la calibrazione dei dati. I suoi interessi scientifici riguardano l'analisi dei getti relativistici di nuclei galattici attivi extragalattici tramite osservazioni interferometriche ad altissima risoluzione angolare, usando strumenti di nuova generazione. Ha pubblicato ad oggi una 80 di articoli su riviste internazionali, partecipato ad altrettanti congressi sia in qualità di invited speaker che di organizzatore scientifico.

Kazi Rygl - [kazi.rygl@inaf.it](mailto:kazi.rygl@inaf.it)

Kazi Rygl e' laureata nel 2006 in Astronomia presso l'Università' di Amsterdam, ed ha conseguito il suo dottorato nel 2010 all'Università' di Bonn. Oggi lavora come ricercatrice all'Istituto di Radioastronomia dell'INAF a Bologna. E' membro dell'Event Horizon Telescope Consortium ed ha contribuito alla pubblicazione nell'aprile scorso della prima immagine del buco nero di M87, partecipando al gruppo di lavoro per la calibrazione dei dati. I suoi interessi scientifici riguardano formazione stellare nella nostra Via Lattea e le galassie vicine tramite osservazione interferometriche ad altissima risoluzione spaziale combinandoli con dati di satelliti ad altre lunghezze d'onda. Ha pubblicato ad oggi circa 90 di articoli su riviste internazionali, e partecipato a molti congressi, sia in qualità di invited speaker che di organizzatore scientifico.



## **Dalla realtà proxy al panico del dasein totale. Il teatro all'epoca dell'AI e dell'AR**

Anna Maria Monteverdi

Università Statale di Milano

**Abstract.** P Può il teatro raccontare le trasformazioni non sempre positive di un mondo sempre più tecnologico? Negli esempi proposti vogliamo portare la risposta di alcuni artisti alla famosa frase di Brecht sul teatro come luogo di riflessione critica e di cambiamento della società stessa.

Hito Steyerl, la più influente voce dell'arte contemporanea, autrice di opere di AI art ha dedicato un libro allo smascheramento della «realtà proxy», della realtà contraffatta o per procura, dell'arte che tenta di sfuggire al «panico del Dasein totale». L'importante oggi è "esserci", presenziare a tutte le ossessive occasioni di aggregazione che internet e l'iperconnessione digitale ci propongono, gestendo il nostro corpo contemporaneamente in più luoghi tramite "proxy", tra avatar digitali e chatbot in una "sovraesposizione dei vissuti" che ci impone la Rete.

**Keywords.** Digital Performance, Media Theatre, AI Theatre, Video mapping

### **Dal Realitysmo al Nuovo Realismo**

Ha appassionato gli spettatori teatrali il recente versante "iperrealtà" della scena: come spiegato nel dossier di "Hystrio" (n.2/2019) sul Teatro della realtà, questo filone di teatro-documentario che ha il suo emblema nelle produzioni del regista Milo Rau, recupererebbe forme e approcci teatrali degli anni Sessanta, espandendosi dalle biografie, all'analisi di fatti cronachistici agli spaccati socio-politici, alle storie di emarginazione sociale. Un ritorno a una realtà "non emendabile" – come afferma il filosofo Maurizio Ferraris nel Manifesto del Nuovo Realismo – contro falsificazioni, mistificazioni mediatiche e omologazioni. Congedo dal realitysmo e approdo al Nuovo Realismo: "Non si può fare a meno del reale, del suo starci di fronte e non essere disponibile a negoziare. Sia quello che sia, ci renda felici o infelici, è qualcosa che resiste e che insiste, ora e sempre, come un fatto che non sopporta di essere ridotto a interpretazione, come un reale che non ha voglia di svaporare in reality" (Ferraris, 2013).

### **Verso una Realtà proxy: l'Arte al tempo del "panico del dasein totale"**

Duty free art (2018) è il titolo del libro dell'artista e filosofa tedesca di origine giapponese Hyto Steyerl, indicata come la più influente artista contemporanea dalla rivista "Art Review"; la Steyerl si occupa di Intelligenza Artificiale: crea opere con uso di AI e riflette sulle trasformazioni (e deviazioni) della società all'epoca di quello che lei definisce, con un neologismo molto efficace, il "circolazionismo" delle immagini. La Steyerl parla di presenza, tema evidentemente molto teatrale: mai come oggi, dice l'artista, il vero valore dell'arte dipende dalla presenza, citando l'ultima significativa performance di Marina Abramovic: *The Artist is Present*.

Oltre a produrre opere, l'artista deve produrre presenza: presenziare a tutte le ossessive

occasioni di aggregazione che Internet e l'iperconnessione digitale ci propongono, tra avatar digitali e chatbot in una "sovrapposizione dei vissuti" che ci impone la Rete, in una vera performance dell'esserci. Il corpo non è essenziale, è essenziale la presenza, che è secondo la Steyerl, una "modalità di investimento". Questa sorta di economia globale della presenza genererebbe, di conseguenza, quello che l'artista definisce un "panico da dasein totale". Ma che realtà producono queste appendici di noi stessi, queste informazioni automatiche gestite da nostri sostituti in forma di assistenti, che mandano tweet al posto nostro, che chattano coi nostri profili? Producono, appunto, dice la Steyerl una realtà proxy, una realtà contraffatta o per procura. Una realtà indotta dal sistema.

### **Dall'illusione della realtà alla Realtà vista dalle macchine**

Verificando quale sia stata la tecnologia più usata dagli artisti in vari ambiti negli ultimi anni, e cioè il sistema A.R. di proiezione architeturale chiamato video mapping, ci si può rendere conto che si è effettivamente andati sempre più verso una richiesta di azione, interazione, presenza di una live technology. Dalla scena urbana si approda a teatro: sempre di più le scenografie sono ricoperte di questa illusoria e neobarocca "luce video mapping" che crea una realtà che non esiste. L'illusione della realtà ha però, un limite legato alla percezione. La tecnica del mapping and masking sfrutta la pre-distorsione dell'immagine o del filmato per farlo apparire non distorto sulla superficie da mappare: la proiezione del contenuto 3D deve essere prima di tutto, perfettamente omografa. Qualunque alterazione della distanza e dell'angolo di incidenza del fascio luminoso implica modificazioni dimensionali e prospettiche dell'immagine, e conseguentemente, irregolarità geometriche e defocalizzazioni più o meno estese.

### **Dall'illusione all'interattività: il videomapping e il motion tracking design**

La fragilità della percezione e dell'illusione, e l'esigenza di avvicinarsi sempre di più alla realtà ha portato a un'evoluzione del mapping: alla ricerca di ciò che determina l'esserci della tecnologia a teatro, gli artisti hanno incontrato l'attore, il performer, il danzatore. Il videomapping statico viene perfezionato con l'aggiunta di sistemi live di interazione tra la proiezione e la figura umana. La Kinect è il controller cinetico di Xbox che entra in commercio nell'ambito dei videogame nel 2010: ha una barra motorizzata in grado di seguire tramite video, i movimenti dei giocatori. La periferica è in grado di rilevare e leggere fino a sei corpi in simultanea. Adrian Mondot (del duo AM&CB) crea proprio con la Kinect, il prototipo di uno spettacolo di danza ad alto tasso di interattività e di illusione della realtà, pur con una tecnologia ap-

Fig. 1

Hyto Steyerl, How not to be seen (2016). La Steyerl elenca in video un bizzarro inventario di soluzioni tecniche per diventare invisibile alle telecamere di sorveglianza "intelligenti".





Fig. 2  
A.M & C.B.,  
Cinématique, 2012

parentemente semplificata: Cinématique (a cui seguiranno altri) si è sviluppato a partire dal software proprietario Emotion che integra dinamica di movimento, oggetti grafici e mapping grazie proprio alla Kinect.

### **Ma non è abbastanza reale! Il live face mapping**

Il live face mapping è un formato che si è andato a perfezionare a partire da un prototipo di Nobumichi Asai del 2014 presentato ad Ars Electronica. Si chiamava Omote, termine giapponese per indicare la maschera (la maschera giapponese Nogaku intesa come zona liminale tra esterno e interno). Venne usato un sistema integrato di face tracking (con uso del sistema di motion capture Optitrack) e projection mapping (con il software Live mapper). Era tecnicamente molto difficile per l'epoca tracciare e proiettare immagini aderenti al volto in tempo reale. Nessuno lo aveva fatto prima: bisognava rilevare i dati dei sensori, elaborarli, calcolare la posizione e l'orientamento del viso, far combaciare il modello in Computer Graphics del viso con la texture animata e inviare l'immagine sul volto tramite proiettore in tempo reale.

### **Ma non è ancora abbastanza reale! Verso il Dynamic Projection Mapping**

La soglia di accettabilità dell'imprecisione nella rilevazione e della latenza è stato il limite e l'asticella da superare in questi anni. Alla fine del 2015 il team di ricerca di Yoshihiro Watanabe nei laboratori Ishikawa dell'Università di Tokyo, sperimenta il Dynafash, un sistema combinato di tracciamento e proiezione in mapping di tipo dinamico (DPM), cioè su un oggetto in movimento e superficie non rigida, con un proiettore ad alta velocità (fino a 1000 frame al secondo a 8 bit con 3 millesimi di secondo di latenza). Le tecnologie di visualizzazione precedenti non erano in grado di fare rilevazioni sufficientemente rapide e questo causava il disallineamento nell'interazione della proiezione dell'immagine con la superficie. L'occhio percepiva l'errore e non accettava di venire "ingannato". Questo procedimento è stato il primo significativo passo verso l'applicazione, anche a teatro, dei sistemi di Computer Vision.

### **La svolta algoritmica: la Realtà vista dalle macchine**

Nel teatro tecnologico tutto cambia con quella che è stata definita da Simone Arcagni "la svolta algoritmica" ovvero "il passaggio all'uso di software di Computer Vision in grado di trasformare in informazioni visive gli input di vari sensori, e dal machine learning che mette a disposizione delle macchine un'ampia autonomia decisionale" (Arcagni 2019).



Fig. 3  
N. Asai, Omote project, 2014. In evidenza i "marker" sul volto della modella che sono la "guida" per il sistema di motion capture Optitrack e per la proiezione in tempo reale delle animazioni in CG tramite Live Mapper

Nel 2017 la Kinect esce di produzione, gli analisti del mercato dicono che è dovuto al fatto che per i videogames oggi è più efficace la Realtà Virtuale. A mio avviso il tema invece, è quello della necessità di un riconoscimento sempre più preciso, "reale" del corpo. Oggi qualunque dispositivo mobile ha la possibilità di effettuare questo riconoscimento. Come si legge dalle caratteristiche tecniche del Face Id di Iphone, la telecamera True Depth utilizzata per l'autenticazione del cellulare della Apple con il volto, acquisisce con precisione i dati, analizzando più di 30.000 punti per creare una mappa di profondità del viso. Il famoso algoritmo di Face detection di Viola-Jones vede la luce già nel 2001 e apre le porte a molteplici sperimentazioni di Computer Vision, aumentando il raggio di applicazioni della Intelligenza Artificiale.

Nel 2009 nasce IImageNet, enorme database di immagini ideato dall'ingegnere cinese Fei-Fei-Li e dal suo staff: è un tentativo di classificare ogni elemento della realtà con un nuovo glossario, attribuendo a ogni immagine non più solo parole chiave, ma una serie di synonym set or synset, cioè una rete di parole associate all'immagine. Una nuova interpretazione del mondo a vantaggio delle macchine che si apprestano a conoscere il mondo e a interpretarlo. Una nuova generazione di sistemi basati sul machine learning e sulle reti neurali sta rendendo possibile rilevare il linguaggio del corpo direttamente dalle immagini: dopo venti anni di sperimentazioni con la motion capture per registrare azioni e imprimerle a un personaggio virtuale, oggi con un sistema come Openpose tutto è diventato automatizzato e semplificato: la telecamera riconosce qualunque tipo di movimento del corpo, gesti e espressioni in tempo reale.

### **Cosa dice la vostra AI di voi?**

Fece scalpore un post scritto su un social network del grande teorico russo dei new media Lev Manovich che provò su di sé un test di AI risultando un "uomo di 38 anni di razza sudamericana". Qualcosa non torna nell'analisi dei dati: eppure questa funzione predittiva (predictive policing) delle AI. è in uso dalla Agenzia Nazionale per la Sicurezza negli States. Nel simposio tenuto da Hyto Steyerl al Castello di Rivoli a dicembre 2018 e in particolar nel suo speech ha parlato delle conseguenze sociali e politiche derivata dall'uso sistematico degli automatismi delle AI per identificare sospetti e introdurre un sistema di controllo sociale.

### **Sight machine-Una questione di sguardi**

Un altro artista sta rivelando il potenziale pericolo insito nella violazione della privacy,



Fig. 4  
Kronos Quartet e Trevor Paglen,  
Sight Machine (2017).

dei dati privati: il tedesco Trevor Paglen, definito per il suo impegno attivista lo “Snowden dell’arte”. La stessa tecnologia di CV Computer Vision usata per le telecamere di sorveglianza di ultima generazione, viene applicata da Paglen a un concerto live con i Kronos Quartet dal titolo significativo Sight machine. Una telecamera riprende i musicisti, riconosce oggetti e volti e interpreta le espressioni dei performer suggerendo ipotesi sull’orientamento sessuale e politico. Le macchine ci osservano e vedono la realtà attraverso i dati che noi stessi abbiamo messo in rete. Ma quali informazioni stiamo dando all’Intelligenza Artificiale?

La stessa telecamera che riprendeva i musicisti viene posizionata nel finale, a riprendere il pubblico: si apre uno scenario inquietante non più solo da 1984 ma da Black mirror. Il teatro, come ben auspicava Brecht, torna a essere luogo per una riflessione critica collettiva della società e di una necessaria riacquisizione della Realtà a cui rischiamo di non avere più accesso diretto. Così Paglen nella scheda dello spettacolo: “When the data we feed the machines reflects the history of our own unequal society, we are, in effect, asking the program to learn our own biases”.

## Riferimenti bibliografici

S. Arcagni, L’occhio della macchina, Torino, Einaudi, 2019.

M. Ferraris, Manifesto del Nuovo Realismo, Bari, Laterza, 2013.

J. Kaplan, Intelligenza artificiale. Guida al prossimo futuro, Roma, Luiss University press, 2018

A.M. Monteverdi, Nuovi media nuovo teatro, Milano, Franco Angeli, 2011

H. Steyerl, Duty free art. L’arte nell’epoca della guerra civile planetaria, Johan Alevi, 2018.

## Autrice



**Anna Maria Monteverdi** - [anna.monteverdi@unimi.it](mailto:anna.monteverdi@unimi.it)

È ricercatrice di Storia del Teatro all’Università Statale di Milano, docente aggregato di Storia della Scenografia. Esperta di Digital Performance è stata docente di Drammaturgia multimediale, Digital video, Storia del Teatro in Università e Accademie di Belle Arti; ha pubblicato: Le arti multimediali digitali (Garzanti 2005), Nuovi media nuovo teatro (2011), Memoria, maschera e macchina nel teatro di Robert Lepage (2018), Leggere lo spettacolo multimediale (2019). Fa parte della rete europea dei teatri EASTAP, dal 2017 è coordinatrice della sezione Intermediality dell’IFTR.

# Connecting document interaction to support music distance learning environments: development of a multipurpose, collaborative PDF reader for real-time sharing of navigation and annotation on sheet music

Enrico Pietrocola

Conservatory of Music G.Verdi of Milan

**Abstract.** This paper presents the planning and development conducted so far regarding the creation of a real-time cooperative application for sharing and visualizing score and textual material, in particular in support of music distance learning and practicing. The application is open source, easily accessible and executable thanks to its predisposition to run on common and widespread devices and operative systems. Interaction between musicians is enhanced by sharing notation making on the score, enabling graphical interaction usually not supported by low latency video conference systems used in music distance learning.

**Keywords.** Interaction, Distance Learning, Music Sheet, PDF, Annotations

## Introduction

Networked Music Performance (NMP) and overall Music making in cooperative environments is spreading quickly. During the last few years we have seen many institutions adopting hardware and software technologies enabling musicians to play music in an almost seamless experience in terms of audio and video.

A dimension that is usually missing from this environment is the complementary dimension of interaction on the score, which is in fact, usually physically shared in a traditional learning context, as for example during a piano lesson, studying side by side with a teacher.

A possible solution to the issue is an application able to display music scores and provide interaction tools, shared between the participants, acting as if the musical document was a single physical entity. In terms of immersion and accessibility, this might prove more effective than expected in improving the feeling of presence between the distance learners.

Features range from brushes to synchronized page turning, layers for orchestras to stack scores in a single document and allow a director to have a single view in which annotations can be set on different groups or instruments. Having an interface that emulates the score's presence provides a key physical feature during the preparatory networked music interaction: pointing on the score, feature that a traditional camera-monitor conference setup cannot provide most of the times.

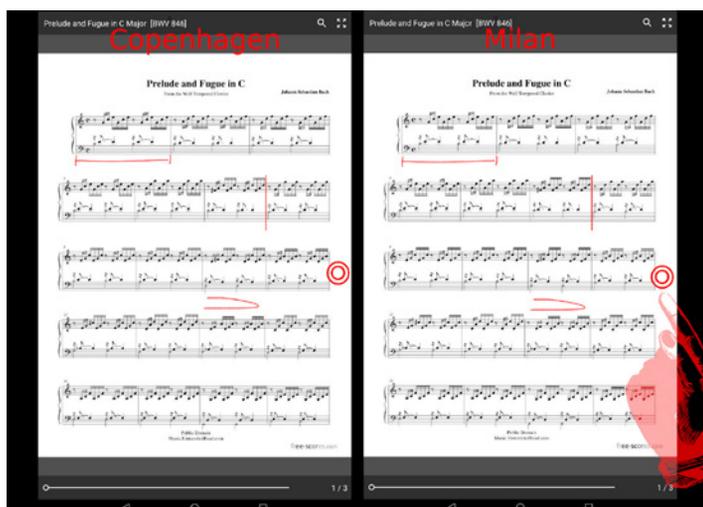


Fig. 1  
Concept example of  
interaction between  
two different app  
users connected at  
realtime

## 1. Implementation

A key role in the implementation of this application comes from the Conservatorio di Musica “Giuseppe Verdi” Milano (ConsMi)<sup>1</sup> and the Intermusic project<sup>2</sup>, acting as hosts in this research and providing the needed material and musicians to test and evaluate the development and design of the app’s first version.

### 1.1 Intermusic

In particular Intermusic is the place in which the idea for this project was born. Named after the concept of “Interactive Environment for Music Learning and Practicing” (Erasmus+ funded 2017 - 2020), aims to create an online shared Platform for that branch of distance learning dedicated to music teaching and practice that will enable modelling and sharing the best training practices for musicians as well as joint courses and online projects.

The two main objectives of Intermusic are:

- support online synchronous singing, instrumental solo performance and chamber groups classes where teachers and students interact in real time in an audio-video communication environment of the highest quality
- use and integrate methodologies and technologies of Blended Learning and Virtual Learning Environments in music higher education<sup>3</sup>.

Therefore, the nature of this project is strongly related with Intermusic and its intellectual outputs.

### 1.2 Guidelines

Development planning followed the guidelines below:

- Open Source: further development after Version 1 will be carried and branched freely by anyone who wants to contribute to the project through a public repository system
- Extendable: being open source will allow developers to create their own versions and most importantly, portings to other platforms. In fact, the first version targets the Android platform, and a Windows version is being planned, but the open source code will

show anyone interested in porting the application how it is structured inside and they will only have to translate it to the new platform's code.

- Multipurpose and modular: the features in the application follow a modular design that enables different uses of the software depending on the circumstances. Users will be able to set up the modalities of interaction, for example:
  1. Performers will be able to open the same score in order to interact on it while rehearsing together.
  2. Directors will be able to open multiple scores at once and create annotations on single parts or groups of both locally and remotely connected performers.
  3. Teachers will be able to interact with students while providing feedback with the help of add-ons like tap tempo metronomes and solfege helper tools.
  4. Lightweight: the application features a very lightweight PDF rendering library, available for many platforms (Windows, MacOS, Android), which is free for open source projects and performs outstandingly well even on low end devices. It was in fact clear since the planning phase that one of the key features in this context would have been performance: avoiding fluidity issues on navigation and overall interaction on the document is vital in order to preserve the feeling of presence.

## 2. Stage of development

The first version of the application is in a pre-alpha stage at the time of writing this article and undergoing further development and design choices. The first version to be published will be on the Android platform since Android tablets are the most diffused and low-price interfaces that will surely enable the widest range of users.

A design and development document will be released with the alpha version and its open source code, its objective is to help create a community of users and developers that will contribute to the growth of the application and its tools.

Before the code and the documents are published, internal tests with the Intermusic team will document interaction and feedback on the application in order to provide use cases and overall aspects of how the application is perceived.

## 3. Future possibilities

This first stage of development is crucial to fix guidelines and concepts for further development. While the main Android branch is expanded with new features and optimizations, it is important to involve more users to use it and join its development, so that the following versions could be shaped after the needs of the community. For this, a forum and a website will be set up all together with all the documentation and code for the beta distribution. Overall, the objective for the future is to support different platforms with cross-platform connectivity and modularity in order to keep the software as open to possibilities as attainable.

## 4. Conclusions

This article showed the idea and main features of an application capable of replicating

human interaction through sheet music on a distance to increase the feeling of presence and improve collaboration in distance learning environments.

## References

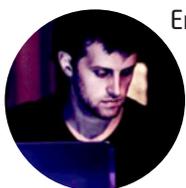
1 <http://www.consmilano.it/>

2 <http://intermusicproject.eu/>

3 <http://intermusicproject.eu/index.php/the-project/content-and-targets/>

The Consortium is composed of the Conservatory of Music “G. Verdi” of Milano (Coordinator), the Politecnico of Milano, the RDAM Royal Danish Academy of Music of Copenhagen, the LMTA Lithuanian Academy of Music and Theatre of Vilnius, the AEC Association Europeenne des Conservatoires, Academies de Musique et Musikhochschulen

## Author



Enrico Pietrocola - [enrico.pietrocola@consmilano.it](mailto:enrico.pietrocola@consmilano.it)

I am an audio artist focused on algorithmic composition and multimedia, with experience in audio programming, sound design and videogame development.

My interests are electroacoustic and noise music, audio-interactive installations and everything involving interaction overall.

In 2018 I won a “Orio Carlini” scholarship from GARR and since then my main objective has been creating a platform and interface to empower interactivity on textual and musical material on a distance.

# MineHEP: Data Mining in High Energy Physics

Laura Redapi<sup>1,5</sup>, Andrea Cioni<sup>2</sup>, Maria Vittoria Garzelli<sup>1,3,4</sup>, Andrea Ceccarelli<sup>2</sup>, Piergiulio Lenzi<sup>1,3</sup>, Vitaliano Ciulli<sup>1,3</sup>

<sup>1</sup>Università degli Studi di Firenze, Dipartimento di Fisica e Astronomia,

<sup>2</sup>Università degli Studi di Firenze, Dipartimento di Matematica e Informatica,

<sup>3</sup>INFN, Firenze, <sup>4</sup>Tuebingen Universitaet, Institute fur Theoretische Physik,

<sup>5</sup>Consortium GARR

**Abstract.** MineHep è un progetto interdisciplinare che nasce dalla collaborazione tra il Dipartimento di Fisica e Astronomia ed il Dipartimento di Matematica e Informatica dell'Università di Firenze, con l'obiettivo di fornire una serie di strumenti software a supporto dell'archiviazione e dell'analisi dei dati raccolti dalla comunità della Fisica delle Alte Energie negli ultimi sessant'anni. L'obiettivo innovativo di MineHep è fornire strumenti per estrarre informazioni rilevanti in modo coerente da un'enorme mole di dati, utilizzando avanzate tecniche di data mining e data warehousing. Queste tecniche, solitamente applicate in campo aziendale, vengono utilizzate qui in Fisica delle particelle per la prima volta. Questo progetto mira a facilitare la libera e pubblica diffusione di dati e dei risultati scientifici a supporto delle infrastrutture di ricerca. Il Consortium GARR ha partecipato al progetto in questa prima fase finanziando una borsa di studio di un anno

**Keywords.** Data Mining, Data Warehousing, Open Data, High-Energy Physics

## Introduzione

Nonostante il successo del Modello Standard (MS), evidenziato dalla scoperta del bosone di Higgs avvenuta nel 2012, i fisici continuano a ritenere che si tratti di una teoria incompleta, in quanto non spiega alcuni aspetti del cosmo (come, ad esempio, la presenza di materia oscura) e non risponde ad alcune questioni concettuali fondamentali (come, ad esempio, la gerarchia tra le masse fermioniche). Sebbene tutte le ricerche esplicite di particelle diverse da quelle contenute nel MS ad oggi non abbiano portato alla scoperta di alcuna particella di Nuova Fisica (D. Kazakov, 2017), la presenza di Nuova Fisica potrebbe essere rimasta invisibile alle singole analisi, ciascuna ristretta ad un sottoinsieme limitato di dati, poiché sparsa tra canali diversi, esperimenti diversi e collider diversi. Al giorno d'oggi, all'interno della comunità della Fisica alle Alte Energie (HEP), non abbiamo modo di gettare uno sguardo di insieme che abbracci contemporaneamente tutti i dati raccolti, da una prospettiva e con uno sforzo di interpretazione globale.

MineHEP si propone come un nuovo approccio per estrarre informazioni dall'intero insieme di dati pubblicati in articoli scientifici dalla comunità di HEP. MineHEP è un cambio di paradigma nella ricerca di Nuova Fisica. Mentre solitamente l'analisi dei dati in HEP è organizzata come studio della compatibilità di un particolare modello di Nuova Fisica con specifici set di dati sperimentali contenuti in un unico documento scientifico, Mi-

neHep beneficia della disponibilità dei dati già presenti nel database HepData (e dei molti altri che vi confluiranno in futuro) per testare le ipotesi in modo globale, consentendo l'analisi simultanea di più dati e la reinterpretazione di vecchie informazioni nel contesto di nuovi modelli, rendendo i dati raccolti più utili nel tempo.

Dimostreremo che le tecnologie esistenti per Data Warehousing (DW) solitamente usate in campo aziendale possono essere portate con successo in nuovi settori, ottenendo risultati innovativi e utili a risolvere diverse criticità.

## 1. Da HepData a MineHEP

### 1.1 Pregi e limiti di HepData/HEPData

HepData (High Energy Physics Database) è stato costruito presso la Durham University in Inghilterra, a partire dagli anni sessanta, e attualmente contiene i dati di migliaia di pubblicazioni, incluse pubblicazioni relative alle collisioni al Large Hadron Collider (LHC). Una revisione portata avanti nel decennio 2000-2010 nell'ambito del progetto CEDAR, ha condotto a una versione intermedia, ancora consultabile pubblicamente (<http://hepdata.cedar.ac.uk>), che prevedeva un database di tipo relazionale (MySQL) e offriva la possibilità, ad un utente esperto, di accedere a quei dati attraverso un'interfaccia web basata su Java. A causa della struttura particolarmente complessa ed articolata, risulta difficile effettuare azioni di data mining su questa versione. Questo è stato uno dei motivi che ha portato ad importanti modifiche a favore di nuove tecnologie e strumenti più adatti alla analisi di dati.

Una nuova versione (<https://www.hepdata.net>), che fa uso di PostgreSQL è stata rilasciata nel 2016. Ad oggi la ricerca su HEPData degli articoli scientifici di interesse, che rispondono a valori fissati di determinati parametri, risulta molto veloce ed efficiente, ma allo stesso tempo, la ricerca è limitata a query sulla base di questi metadati, che producono come output liste di articoli scientifici per i quali i metadati hanno i valori richiesti dall'utente. Risulta quindi possibile una ricerca per articoli, ma non per dati interni agli articoli stessi, limitando anche la possibilità di visualizzare contemporaneamente in un medesimo grafico i dati provenienti da analisi diverse per eventuali confronti. Per rispondere alle nostre esigenze di ricerca e analisi, abbiamo identificato e selezionato tecniche di analisi Online Analytical Processing (Kimball Group website) che sono oggi strumenti consolidati per logiche decisionali di azienda, e le abbiamo esportate e applicate al nostro problema in ambito HEP.

### 1.2 Estrazione di dati e pulitura del database HepData

I membri del gruppo HEP dell'Università di Durham ci hanno fornito un dump del database MySQL che contiene tutti i dati già presenti in HepData fino ad Aprile 2018.

Questa versione ci ha permesso di accedere allo schema di tabelle in cui i dati sono organizzati e di comprendere la relazione tra le varie entità del database, indispensabile per gli steps successivi. Sono state fatte diverse operazioni tra cui l'eliminazione di tabelle vuote e chiavi duplicate, l'eliminazione di alcune informazioni ridondanti, oltre a una prima riorganizzazione dei dati che ha evidenziato l'importanza di rendere omogenei i loro formati.

### 1.3 Progettazione e popolamento del nuovo database MineHep

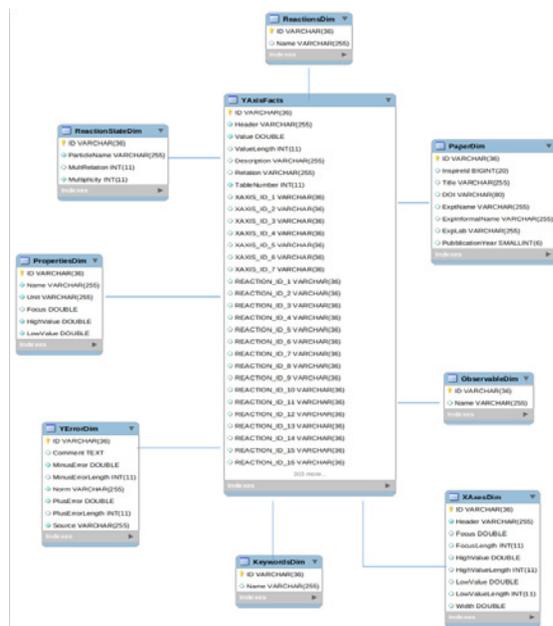
Il modello dimensionale risponde a due requisiti fondamentali per la presentazione di dati analitici: fornire l'informazione in maniera chiara e permettere all'utente di accedervi velocemente. La semplicità assicura che gli utenti possano facilmente comprendere i dati, ma anche che i software possano navigare e consegnare i risultati velocemente ed efficientemente. Una volta pulito il database originale, abbiamo progettato uno schema a stella del tipo mostrato in Figura 1. In uno schema a stella, la tabella dei fatti è l'unica a possedere collegamenti multipli con le altre tabelle, che si concretizzano attraverso gli identificatori univoci (campi chiavi) in essa contenuti. Le dimensioni hanno tutte un unico collegamento alla tabella centrale, con lo scopo di minimizzare il numero di istruzioni richieste per ciascuna query e facilitare l'interrogazione del database da parte di un utente.

Del nostro schema si nota la centralità della tabella dei fatti, YAxisFacts, che contiene i dati misurati di ogni esperimento e, intorno a questa, otto dimensioni che caratterizzano il singolo dato numerico.

Una volta progettato il nuovo database, l'intera procedura di importazione è stata automatizzata attraverso l'uso di Docker e Docker Compose prevedendo l'integrazione con sistemi di generazione dashboard e altri software.

Sono stati infine scritti ed eseguiti manualmente 112 test per verificare che il nuovo database contenesse informazioni coerenti con quelle contenute nel database HepData originale. I test effettuati hanno dato esito positivo.

Fig.1  
Diagramma a stella del  
database MineHep



## 2. Sviluppo della query interface/dashboard per la visualizzazione dei dati di MineHep

Una prima esigenza dell'utente è suddividere in diverse serie i risultati ottenuti e inserirli in un unico grafico per poter fare dei confronti. Il software di DW scelto per la costruzione di

una dashboard accessibile via web, per permettere una prima presentazione dei risultati, è stato Metabase (Metabase website). Metabase è uno strumento open source che permette di interrogare ed estrarre conoscenza dai dati in modo semplice, fornendo un'interfaccia SQL. Abbiamo preimpostato in Metabase delle query in linguaggio SQL, che l'utente può completare attraverso semplice menu a tendina. All'utente è inoltre possibile salvare le queries in una dashboard, dove confrontare risultati. Di seguito riportiamo un esempio per l'esplorazione e il confronto di dati estratti da articoli scientifici diversi:

Cerchiamo tutti i dati sperimentali relativi a distribuzioni differenziali  $d\sigma/dp_T$  per la reazione  $pp \rightarrow B^+ + X$ , con produzione di mesoni  $B^+$  di impulso trasversale nell'intervallo compreso tra 0 e 250 GeV e rapidità  $< 5$ , a fissata energia di collisione, e riportiamoli in un unico grafico.

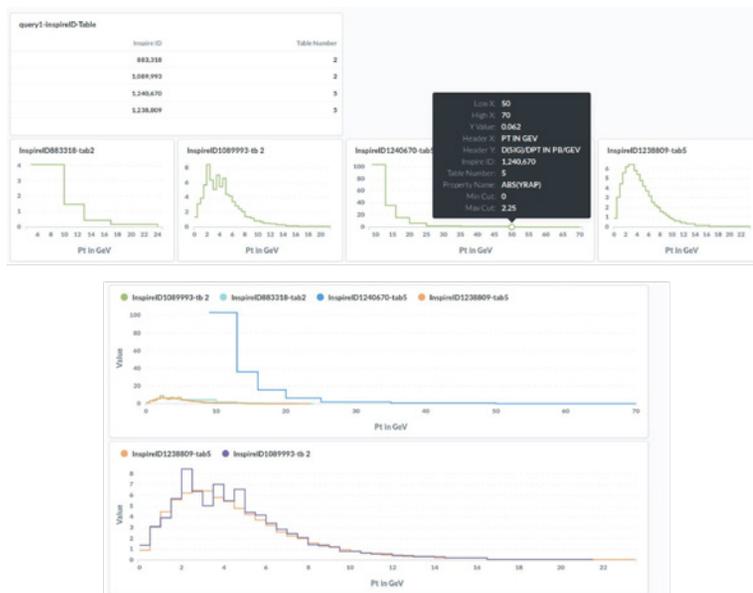


Fig.2  
Output di Metabase alla query di interesse riportata nel testo

Usando Metabase la risposta cercata può essere ottenuta dall'esecuzione di più query in serie: la prima seleziona i papers e le tabelle in cui i dati relativi all'osservabile di interesse sono riportati, mentre le successive estraggono l'informazione, che può essere visualizzata tramite tabella o tramite plot. Infine, è possibile sovrapporre i grafici ottenuti, permettendo di evidenziare differenze o analogie tra i vari set di misure in esame. L'output di Metabase alla successione di query descritta è riportato in Figura 2.

### 3. Conclusioni

Abbiamo dimostrato come la riorganizzazione del database HepData nel database MineHep tramite tecniche di DW, permetta il confronto automatico tra serie di dati sperimentali presentate in articoli scientifici differenti, ottenute sotto le medesime condizioni di analisi. Questa nuova potenzialità può permettere di evidenziare in maniera semi-automatica anomalie di specifici set di dati, che potrebbero indicare la presenza di Nuova

Fisica. L'estensione di tecniche di DW alla riorganizzazione di database di interesse in altri campi scientifici è possibile e può permettere un uso più proficuo dell'informazione già raccolta, migliorando la comprensione di grandi quantità di dati.

## Riferimenti bibliografici

D. Kazakov, "Beyond the Standard Model '17", lectures at the European School of High-Energy Physics 2017, CERN Yellow Reports School Proc. 3 (2018), 83-131

Kimball Group - Star Schemas and OLAP Cubes, <https://bit.ly/2Fff4AY>

Metabase website URL: <https://metabase.com/>

## Autori



**Laura Redapi** - [laura.redapi@gmail.com](mailto:laura.redapi@gmail.com)

Laura Redapi è specializzanda in Fisica medica presso l'Università degli Studi di Firenze. Ha vinto una borsa GARR per l'anno 2018/2019 durante il quale si è occupata del progetto MineHep estendendo l'utilizzo delle tecniche di DW anche a database medici.

**Andrea Cioni** - [andrea.cioni6@stud.unifi.it](mailto:andrea.cioni6@stud.unifi.it)

Andrea Cioni si è laureato in Informatica presso l'Università degli Studi di Firenze nel 2019, discutendo una tesi sull'uso di tecniche di DW per lo sviluppo del progetto MineHEP. Attualmente è software developer presso Florence Consulting Group.

**Maria Vittoria Garzelli** - [maria.vittoria.garzelli@cern.ch](mailto:maria.vittoria.garzelli@cern.ch)

Maria Vittoria Garzelli è borsista post-dottorato nell'ambito del progetto MineHep, esperta in vari aspetti di fenomenologia delle particelle elementari e nello sviluppo di software per produrre predizioni teoriche da confrontare con i dati sperimentali.

**Andrea Ceccarelli** - [andrea.ceccarelli@unifi.it](mailto:andrea.ceccarelli@unifi.it)

Andrea Ceccarelli è Ricercatore a Tempo Determinato (RTDb) in Informatica presso l'Università degli Studi di Firenze. I suoi interessi principali di ricerca riguardano la progettazione e validazione di sistemi informatici disponibili e sicuri.

**Piergiulio Lenzi** - [piergiulio.lenzi@unifi.it](mailto:piergiulio.lenzi@unifi.it)

Piergiulio Lenzi è Ricercatore a Tempo Determinato (RTDb) in Fisica presso l'Università degli Studi di Firenze. Partecipa all'esperimento CMS a LHC ed è esperto nelle tecniche statistiche di analisi dati agli acceleratori. È Principal Investigator del progetto MineHEP presso l'Università di Firenze.

**Vitaliano Ciulli** - [vitaliano.ciulli@unifi.it](mailto:vitaliano.ciulli@unifi.it)

Vitaliano Ciulli è Professore Associato in Fisica presso l'Università degli Studi di Firenze e Vicedirettore del Dipartimento di Fisica ed Astronomia. Partecipa all'esperimento CMS a LHC per il quale è stato coordinatore del gruppo dei generatori di eventi Monte Carlo.

# MeDB: a Portable Clinical Historical Database

Paolo Ricciuti, Alba Di Pardo, Daniele Cafolla

IRCCS Neuromed

**Abstract.** In medical structures it is very important to keep track of patient conditions in every moment: every exam they come to do, their pathologies and exemptions and their family condition. It is also important for them to have access to their medical history particularly for neurological diseases. This system proposes a web portal in conjunction with a MySQL database and an ad hoc hardware system to facilitate both the data acquisition and the availability of those data for the patient everywhere

**Keywords.** IoT, Open Science, FAIR data & services, Biomedical data research, Portable database

## Introduction

Telemedicine and teleconsultation, which offer potential improved healthcare, are becoming possible by the improvement of IoT technologies. Design and creation of multimedia information management and medical communication technologies has been underway in several studies [Krishnan et al., 1997 Ingman, M., & Gyllensten 2006; Fincke et al., 2010]. The objective of these field has been to develop systems which provide the medical community with the necessary storage, processing and communication facilities. One of the key issues is the uniform and user-friendly handling of multimedia information.

## 1. Concept

The proposed system is composed of three main modules: a database, a web portal and an Arduino-based hardware. The MySQL Database module is responsible of data persistency. It is composed of ten tables, some of them contains lists of data used by the program to populate dropdown and, some of them contains users' information and the remaining are relationship tables created to model "many to many" relationship between records.

The Web Portal module, Figure 1, use a mixture of HTML, JavaScript and PHP. It connects to the MySQL module allowing the user to exec basic CRUD (Create, Read, Update, Delete) operations on tables, like adding or removing an User, an Exam, an Exemption, a Pathology or a Mutation. Through the portal it is also possible to add and remove the relation between those records in a fast and user-friendly way. It is possible to retrieve the last access of every patient for a specific exam and a specific date, to get all the first accesses in a specific time period, to know how many patients have a specific exemption code, how many patients have a specific disease and how many of them have a specific mutation. The latter provides a useful tool that automatically calculate the average arising age giving the researcher an "on the fly" information about a specific mutation.

Finally, an innovative tool that let rapidly create and assign a genetic pedigree to each user

Fig. 1

Web Portal module:

- a) MySQL database,
- b) genetic pedigree tree editor,
- c) Hardware-Portal communication console



of the system, is implemented and made accessible from the user card. This tree editor is a lightweight web-application written using a mixture of vanilla JavaScript and “p5.js” that let exporting the tree as a jpg or save it into the system to be modified later.

The Hardware module, Figure 2, is an Arduino-based system should be installed at the entrance of every exam room or in a medical doctor office. The same system is also used for the registration phase, in which an USB card is associated in a “one to one” relation with an User. After the registration, every time a user inserts his USB card in the device, it communicates with a central server through Wi-Fi; the server receives the ID of the patient and the ID of the exam and register a new access into the Database. It is important to empathize that this kind of system gives the medical structure the freedom to register patients by his own. This help categorize patients from different medical departments giving the possibility to collect data only from their field of interest.

Furthermore, the server sends updated personal and clinical information to the Hardware module, these data will be stored in the personal card. The advantage of such a system will be that the patient will have always the data with him and only medical doctors can access and modify the database. This will avoid the loss of information going through different medical doctors and environments and can help them to treat patients without a doubt about patient medical history in case of emergency.

## 2. Technologies

The proposed system has its foundation in the hearth of several main technologies: MySQL, PHP, HTML, JavaScript, C# and Arduino. These technologies have been chosen to create a product that is solid, fast, easy to use, install and maintain.

MySQL is a golden well-known standard in the field of databases and it has enormous capability to store and manipulate data.

PHP has been chosen to write the Web Portal module. This programming language is very fast and has a lot of built in functionalities to connect and execute query on a Database. HTML, the main language used to build the web portal. Php is a language that works in couple with HTML, it can be integrated inside of it using variables and database calls to create a dynamic web page. Furthermore, JavaScript has been used to create a more dynamic page: this programming language can act on the DOM (Document Object

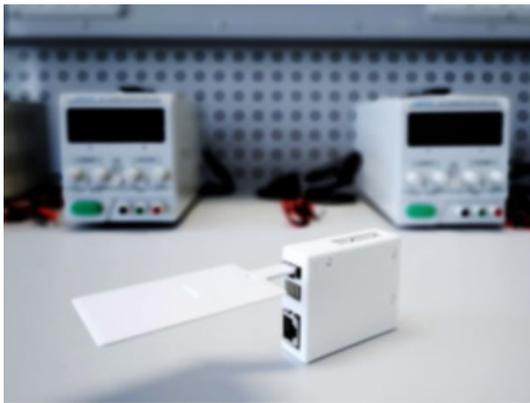


Fig. 2  
Hardware module  
with ID Card

Model) elements of the pages changing their styles, their position or even injecting text inside them. In this web portal it is used to create a more user-friendly GUI (Graphic User Interface) in conjunction with CSS (cascade style sheet).

Complementarily C# has been used since it is a very powerful object-oriented programming language and make the system to be portable and fast. This language is used to build the server that will receive and dispatch information from the Arduino.

The hardware module for the identification of the patient and the storage of personal and clinical data consists of an Arduino YÚN together with an USB shield programmed using the specific Arduino Environment.

Arduino YÚN has been chosen since it is strictly designed for IoT world. It combines the power of Linux distribution based on OpenWrt with the ease of use of Arduino environment. Furthermore, the board has built-in Ethernet and Wi-Fi support. Finally, the system can be powered by plugging it with a plug or simply connecting it to a battery.

The developed software uses a conjunction of a series of libraries: “SPI.h”, “Usb.h”, “Process.h”, “Bridge.h” and “BridgeClient.h”. Those libraries are used by Arduino to access the USB card and read the vendor ID: a unique ID assigned from the factory to each pen. Furthermore, those libraries are used to create a Client to connect to the Server that will later associate this ID to an account on the Web Portal. If necessary, the hardware module can download updated information to the USB card that can be not only textual information but also medical images coming from medical devices.

### 3. Conclusions

The proposed solution aims to create an all-in-one system that is portable, easy to use and install, low-cost and that can be helpful for patients, medical doctor and researchers. It allows to easily access, manage and store crucial information about patients’ medical history from everywhere letting researchers and medical doctors create ad-hoc optimum therapies. Furthermore, it permits statistical analysis on data helping the medical centres to focus their efforts on specific pathologies and needs.

### References

Fincke, B. G., Miller, D. R., & Turpin, R. (2010). A classification of diabetic foot infections

using ICD-9-CM codes: application to a large computerized medical database. *BMC health services research*, 10(1), 192.

Ingman, M., & Gyllensten, U. (2006), mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences, *Nucleic acids research*, 34(suppl\_1), D749-D751.

Krishnan, S. M., Goh, P. M. Y., & Datian, Y. (1997), A multimedia-based medical database network system for special clinical procedures in healthcare delivery, In *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 'Magnificent Milestones and Emerging Opportunities in Medical Engineering'* (Cat. No. 97CH36136), Vol. 2, pp. 936-938.

## Authors



**Paolo Ricciuti** - [ricciutipaolo@gmail.com](mailto:ricciutipaolo@gmail.com)

He got a degree in Computer Science at the University of Molise in 2009 with a thesis "Design and Development of a Virtual Darkroom". In 2017 he started the master degree in Computer Science at Federico II university in Naples, focused on Artificial Intelligence. In 2018 he started a stage period at "NTTData Italia S.p.A" and in 2019 it started working for the "Biomechatronics Lab" at IRCCS Neuromed, focusing on signal processing, software development to support research and clinical needs.

**Alba Di Pardo** - [alba.dipardo@neuromed.it](mailto:alba.dipardo@neuromed.it)

She is a geneticist with experience in the field of genetic rare diseases and neurodegeneration. She has more than 15 years of research experience and of professional genetic testing and counselling for genetic conditions including Huntington Disease. Presently, she works at Centre for Neurogenetics and Rare Disease at IRCCS Neuromed, Italy. She has extensive experience in preclinical and clinical research, genetic testing and counselling for patients affected with neurological rare diseases.



**Daniele Cafolla** - [contact@danielecafolla.eu](mailto:contact@danielecafolla.eu), [bioingegneria@neuromed.it](mailto:bioingegneria@neuromed.it)

He got the Dual MSc in Mechanical Engineering in 2012 at Cassino, Italy and Panamerican (Mexico) Universities. From October 2013 to May 2014 he developed a Human-like Torso at NTU, Singapore. After his PhD, he worked in several fields as: Biomechanics; Humanoid, Exploration, Rehabilitation robotics; Rapid prototyping. Since 2018 he is the head of Biomechatronics Lab at IRCCS Neuromed, working in: Biomechanics; Humanoid, Exploration, and Rehabilitation robotics; Rapid prototyping.

## Banche dati di documenti giuridici per la formazione e le digital humanities

Francesco Romano, Antonio Cammelli

Consiglio Nazionale delle Ricerche, Istituto di Informatica Giuridica e Sistemi Giudiziari

**Abstract.** La relazione rende conto dell'infrastruttura di ricerca costituita da alcune banche dati del CNR messe a disposizione dall'Istituto di Informatica Giuridica e Sistemi Giudiziari (ex ITTIG) sul proprio sito ([www.igsg.cnr.it](http://www.igsg.cnr.it)). Queste banche dati rappresentano un'infrastruttura di ricerca per la comunità scientifica sia nazionale che internazionale, per gli studiosi della lingua, per gli operatori del diritto, ma anche per chi vuole comprendere come cambia la nostra società

**Keywords.** Linguaggio giuridico, diffusione informazione giuridica, banche dati giuridiche, portali pubblici

### Introduzione

L'elaborazione, l'utilizzo e lo sviluppo di strumenti tecnologici nel campo delle scienze umane (dalla storia alla filologia, dall'archivistica alla storia dell'arte) è ciò che ormai da alcuni anni è inquadrato con la definizione di Digital Humanities. Si può parlare di vera e propria disciplina alla quale sono dedicati corsi di studio, Master, convegni, pubblicazioni. Tuttavia l'incrocio tra informatica e scienze umane e sociali è fonte di criticità.

Come già notano alcuni autori: passare "dall'esame di un testo, o di molti testi", fatta dall'occhio e dal "cervello di un lettore", alla scansione automatica fatta da un "delegato tecnologico, può portare in alcuni casi a risultati interessanti". Però è altrettanto vero che un "approccio digitale, a fatti non solo quantitativamente ricchi ma anche concettualmente complessi", può produrre analisi superficiali di testi che erano stati concepiti per essere letti, interpretati e valutati dall'occhio (e dal cervello) umano (Tomasin 2017: 46). Insomma l'uso delle tecnologie applicate ad alcune materie del sapere rischia di comportare una sorta di gamification di tali discipline, anche se biblioteche, archivi, collezioni, potranno essere sempre più "supportati dalla tecnologia", qualora essa rimanga funzionale e accessoria al valore di prodotti di una "cultura dotata di autonoma dignità" (Tomasin 2017: 138).

In tale ottica le banche dati di documenti giuridici antichi e contemporanei disponibili sul sito dell'Istituto di Informatica Giuridica e Sistemi Giudiziari (ex ITTIG) del CNR costituiscono un'infrastruttura utile per studiare l'evoluzione del linguaggio giuridico italiano e non solo di esso, perché uniscono alle caratteristiche proprie dell'elaboratore informatico, la possibilità di visualizzare e quindi analizzare il risultato della ricerca.

## 1. Banche dati per le digital humanities

Oltre alle banche dati LGI (Lessico Giuridico Italiano) e LLI (Lingua legislativa Italiana) contenenti la prima più di 900.000 schede-immagine usate per gli spogli del Vocabolario Giuridico Italiano, la seconda oltre 200 testi integrali di costituzioni e codici antichi e moderni, il sito IGSG offre la possibilità di consultare preziosi documenti quali le gride e gli editti dello Stato di Milano (documenti dal 1560 al 1796) e la legislazione medicea nelle raccolte dell'Archivio di Stato di Firenze (copertura dal 1532 al 1737), mentre una prossima tappa del percorso di condivisione digitale vedrà mettere in linea i contenuti di importanti manoscritti giuridici che attualmente sono digitalizzati solo su CD. In tali archivi e banche dati sono infatti presenti testi che aiutano a comprendere le diverse epoche storiche del diritto italiano e, più in generale, costituiscono uno strumento per documentare, attraverso varie fonti del diritto, la storia del nostro Paese partendo dal secolo X per arrivare ai nostri giorni. Inoltre i ricercatori del CNR, con l'ausilio di altri studiosi, per mezzo di un software per il lavoro collaborativo in Rete, hanno elaborato uno strumento per una migliore fruizione di tale patrimonio documentario. Ci riferiamo alla banca dati IS-LeGI .

### 1.1 L'Indice Semantico per il Lessico Giuridico Italiano

La banca dati IS-LeGI è implementata grazie a un software che permette di compilare, per un dato lemma il significato che la parola assume nel contesto dato. In tal modo a una certa parola vengono attribuiti rilevanti metadati quali le accezioni e la fraseologia presente nel contesto (scheda-immagine). Così il linguista e il filologo potranno apprezzare le infinite sfumature del lessico giuridico che variano col passare del tempo (Romano et al. 2018). Tali risorse digitali potranno fornire un supporto a chi voglia studiare l'evolversi della nostra società attraverso i vari significati dati nel tempo alle parole: così, ad esempio, è stato verificato l'uso che nel linguaggio giuridico si fa e si è fatto di parole molto discusse quali identità o razza (Cammelli et alii 2019), termini questi ultimi tornati di grande attualità specialmente nel momento in cui il nostro Paese è interessato dal fenomeno migratorio. Inoltre la banca dati IS-LeGI per la sua peculiare caratteristica di essere implementabile per mezzo di un software che permette di agire in modalità cooperativa in Rete, si presta anche ad un uso didattico e formativo, come verificato durante le lezioni del Master La lingua del diritto, dove il software IS-Legi è stato usato in un laboratorio nel quale gli studenti hanno interagito nella redazione di alcune voci della banca dati.

## 2. Banche dati per l'accessibilità del diritto

L'Istituto IGSG rende disponibili anche altre importanti banche dati per l'accesso e la conoscibilità della dottrina giuridica e del diritto positivo. Ci riferiamo alla banca dati DoGi e al Portale PAeSI. La prima banca dati offre abstract di articoli pubblicati in riviste giuridiche italiane. La banca dati è implementata dal 1970 ed il corpus di spoglio è costituito da circa 250 periodici editi in Italia. La banca dati è aggiornata ogni due mesi e mediamente vede un incremento annuo di 13.000 unità documentarie (Conti et alii 2017). Il Portale dell'immigrazione della Regione Toscana – PAeSI – è gestito da IGSG anche con la collaborazione della Prefettura di Firenze. Il portale offre vari servizi tra cui una banca dati di atti

di legislazione italiana del settore immigrazione. Un'ulteriore banca dati raccoglie schede informative sui procedimenti amministrativi in materia di immigrazione ed alcune fra esse, sono state semplificate nel linguaggio, nella struttura sintattica e nell'organizzazione dell'informazione grazie anche all'aiuto dei diversi stakeholder della materia e degli utenti (Fioravanti et alii 2018).

### 3. Conclusioni

Abbiamo evidenziato come le banche dati di IGSG rappresentino un'infrastruttura di ricerca per la comunità scientifica sia nazionale che internazionale (per gli studiosi della lingua, per gli operatori del diritto), ma anche per chi vuole comprendere come cambia la nostra società, usando come parametro di riferimento l'evolversi di un linguaggio specialistico che si affida anche alla lingua comune e che dunque testimonia nel suo lessico e nella sua sintassi le trasformazioni della nostra comunità. Inoltre abbiamo evidenziato come le modalità di implementazione della banca dati IS-LeGI siano funzionali alle attività di formazione. Ma questo tipo di supporto può essere utile anche nell'ambito delle scienze storiche. Se infatti queste ultime sono sempre più messe in discussione, come pare essere evidenziato dall'odierno dibattito scaturito dall'abolizione della prova di storia nell'esame di maturità, è anche vero che la "storia è ormai una questione pubblica, che coinvolge gruppi e comunità, associazioni e capillari reti culturali" e che la "conoscenza storica e la sua pratica non possono limitarsi al mondo accademico e scientifico" ma debbono aprirsi alla collaborazione con "storici, professionali e non" (Ridolfi 2017: 5) anche per mezzo di quella digital public history, che può divenire una nuova "grammatica storica" per le generazioni dei "nativi digitali" (Ridolfi 2017: 24).

### Riferimenti bibliografici

A. Cammelli, C. Fioravanti, F. Romano (2019), La parola "razza": analisi diacronica nei testi giuridici antichi e moderni, in *Rivista di sociologia del diritto*, n. 2, pp. 191-200.

S. Conti, S. Faro, E. Marinai, G. Peruginelli (2017), La banca dati DoGi: struttura, caratteristiche e prospettive, in Sebastiano Faro, Ginevra Peruginelli, "La dottrina giuridica e la sua diffusione", Giappichelli, Torino.

C. Fioravanti, F. Romano (2018), Stakeholders Engagement to Simplify Communication of Administrative Procedures in the Field of Immigration: Experimentation of Methods and Tools, in Shefali Virkar, Peter Parycek, Noella Edelmann, Olivier Glassey, Marijn Jansen, Hans Jochen Scholl, Efthimios Tambouris (Editors) *Proceedings of the International Conference EGOV-CeDEM-ePart 2018*, Edition Donau-Universität Krems.

M. Ridolfi (2017), *Verso la public history. Fare e raccontare storia nel tempo presente*, Pacini editore, Pisa.

F. Romano, A. Cammelli (2018), Strumenti e metodi per lo studio della lingua del diritto. Le attestazioni del termine mandato nei documenti giuridici antichi e contemporanei, in Luca D'Onghia e Lorenzo Tomasin (a cura di), *Etimologia e storia delle parole*, Franco Cesati Editore, Firenze.

L. Tomasin (2017), L'impronta digitale. Cultura umanistica e tecnologia, Carocci Editore, Milano.

<http://www.ittig.cnr.it/BancheDatiGuide/vgi/islegi/index.php>

<http://www.ittig.cnr.it/dogi>

<http://www2.immigrazione.regione.toscana.it/home>

## **Autori**



**Francesco Romano** - [francesco.romano@igsg.cnr.it](mailto:francesco.romano@igsg.cnr.it)

Laureato nel febbraio del 1996 presso la Facoltà di Giurisprudenza di Firenze. Dottore di Ricerca in Telematica e Società dell'Informazione presso l'Università di Firenze. Attualmente è ricercatore presso CNR IGSG.

**Antonio Cammelli** - [antonio.cammelli@igsg.cnr.it](mailto:antonio.cammelli@igsg.cnr.it)

Dirigente di ricerca CNR in quiescenza, attualmente associato alle ricerche di CNR IGSG con particolare riferimento alla gestione e produzione della banca dati storica IS-LeGI.



## Start@unito underground map for an e-learning trip

Marina Marchisio<sup>1</sup>, Sergio Rabellino<sup>2</sup>, Matteo Sacchet<sup>3</sup>, Daniela Salusso<sup>4</sup>

<sup>1</sup>Department of Molecular Biotechnology and Health Sciences, University of Turin, <sup>2</sup>Department of Computer Science, University of Turin, <sup>3</sup>Department of Mathematics "G. Peano", University of Turin, <sup>4</sup>Department of Foreign Languages and Literatures and Modern Cultures, University of Turin

**Abstract.** Start@unito is a project of the University of Turin, born to help students get a head start in their university studies, orientate themselves, and have an overview of university courses. Beyond these objectives, the project aims at reducing the drop-out rate and disseminate the use of digital technologies in didactics. In this paper, we provide a guidance system designed as an "underground map", discussing different possible paths according to the kind of user: students, professors, staff,... and the various kind of tickets that allow local transport in this city

**Keywords.** Digital Education, Online platform, Start@unito, Underground map, Tertiary Education

### Introduction

E-learning is widely spreading, providing fertile ground for research (Duval, Sharples, & Sutherland, 2017). The way people interact with their education is significantly changing too: the learners are more involved and more aware of their study, fitting study with their everyday's life (Norman, 2016). The University of Turin has been working on this topic since the beginning of the e-learning era. Many current projects are centred around this topic, like Orient@mente (Barana, Marchisio, et al., 2017), helping students to prepare for admission tests, Scuola dei Compiti (Barana, Fioravera, Marchisio, & Rabellino, 2017), supporting 8th and 9th grade students with low marks on important disciplines, Problem Posing and Solving (Barana et al., 2018), assisting high school teachers in using technology in their class. In this frame, start@unito is a project created to boost careers in university programs, by accompanying students before and during their studies. It offers several open and free online courses, which can be used for different goals, such as taking exams in advance, for student orientation purposes, finding support materials or even simply just studying online, in particular for students with special needs or people interested in broadening their knowledge. After attending the online courses, students can obtain a certificate, with which they can obtain corresponding ECTS after sitting an in-person exam, which is compulsory for Italian universities.

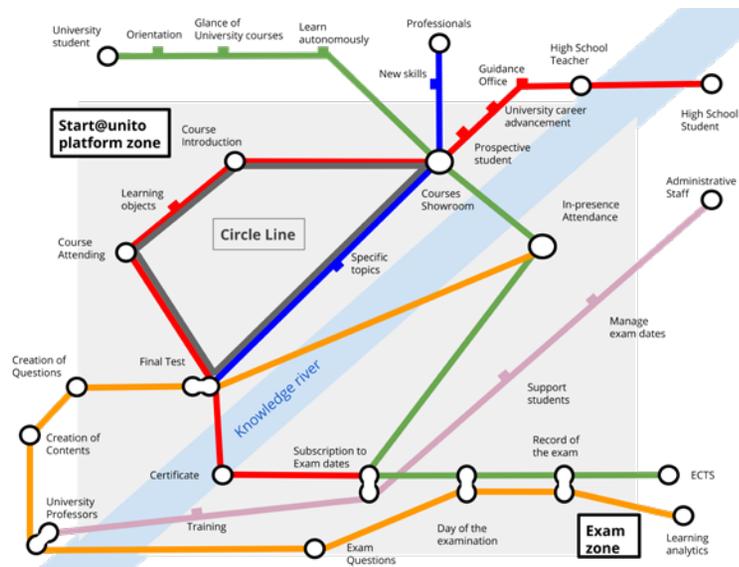
The correct functioning of the project requires a connection between many actors: platform providers, university offices, professors, tutors, users, students. Using a metaphor, each of these actors/citizens has a precise path to follow in start@unito city. When you

visit a new city, you have to find out bus routes, timetables and figure out which way is the fastest, which is usually the underground. In the same way, in order to move around in the start@unito city, we have built an underground map to show you the connections between the nodes. In this work, we will analyze the map.

## 1. Start@unito underground map

Different kinds of people interact with the underground network provided by the start@unito project: the common goal for all users is to improve their learning and broaden their horizons. But every complex system requires a handbook to navigate it. The development of this underground net is neverending work in progress, both from the point of view of didactical contents and from the point of view of ICT. The main lines of the tube map are described in the following paragraphs.

Fig. 1  
Start@unito  
"underground map"



### 1.1 Red line

The end of the line stop is on one side High school students, who are supported by the next stop, High school teachers. These future students usually interact with orientation offices, and they will enter the platform zone with the main aim of starting university in advance. The first page they should visit is the Courses showroom, showing introductory descriptions, videos and the main details about the courses. After attending the online course, with all its learning objects, students may try a final test to evaluate their newly acquired competences and get a certificate, mandatory in order to attend the in-person exam. They need to register for the exam dates and from this point they follow the green line.

### 1.2 Green line

The end of the line stop is on one side university students. They start the trip with different motivations: orientate themselves in the university training environment, learn autonomously, start their studies in advance and follow a personalized path. After watching

the showroom, they can jump off the green line and hop on the red one and attend online courses. Alternatively, they can attend in-person lessons, taking advantage of the online contents, too. After a while, the green line intersects with the red line anyway. Thus, the stops are planned for taking the exam and, after it has been recorded, obtaining the corresponding ECTS, the last stop of their trip.

### **1.3 Blue line**

The end of the line stop is on one side professionals. They usually want to interact with university courses to develop specific skills, or to enhance the quality of personal work. Usually they need to study specific topics, that's why their path to the final test is shorter than the red line. They may even stop there, because the new skill is acquired, but in order to obtain a formal validation, they get on the red line as regular students.

### **1.4 Yellow line**

The end of the line stop is on one side university professors. They are the carriers of knowledge and are essential for the validation of contents. They are also responsible for the creation of contents, interactive materials and questions and the preparation of the final test. They connect this track with their in-person didactics. This is one possible way. On the other side, they are responsible for the examination, prepare computer-based questions, give written or oral examinations. The line ends with learning analytics, studied in order to improve learner experience. The common stations with the students' lines are separated, in the sense that they are connected but you cannot directly jump off and hop on another line, because the yellow one is a lane.

### **1.5 Purple line**

The end of the line stop is on one side the administrative staff. This is again a lane. The administrative staff support students in their career in many ways, in particular, referring to this project, by managing exam dates and support students. The administrative staff reaches also the professors stop, providing them with training on digital education.

### **1.6 Special lines**

The circle line allows user to travel around the platform multiple times and attend more than one course.

### **1.7 Tickets**

Platform managers have a free ticket for the entire "start@unito platform" zone. They can jump on and off lines in this zone whenever they want, even on and off lanes. Red and blue lines are free, while the green line is the only one for which a ticket is mandatory. Users can buy single or multiple tickets.

## **2. Conclusions**

In the present paper we outlined the intersections between the various paths in order to

guide all start@unito users (such as prospective students, teachers, and general citizens) through the necessary steps to improve learning and academic results. The University of Turin is the company that manages this network, which can be shared by other partners such as Italian and foreign universities, since the platform speaks not only Italian but also English (the start@unito city is bilingual) and it is possible to learn new languages (start@unito city is cosmopolitan). The authors invite you to visit the city “always on the move”!

## References

- Barana, A., Brancaccio, A., Esposito, M., Fioravera, M., Fissore, C., Marchisio, M., & Rabelino, S. (2018). Online Asynchronous Collaboration for Enhancing Teacher Professional Knowledges and Competences. The 14th International Scientific Conference ELearning and Software for Education, 167–175. <https://doi.org/10.12753/2066-026x-18-023>
- Barana, A., Fioravera, M., Marchisio, M., & Rabellino, S. (2017). Adaptive Teaching Supported by ICTs to Reduce the School Failure in the Project “Scuola Dei Compiti”. Proceedings of 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), 432–437. <https://doi.org/10.1109/COMPSAC.2017.44>
- Barana, A., Marchisio, M., Bogino, A., Operti, L., Floris, F., Fioravera, M., & Rabellino, S. (2017). Self-Paced Approach in Synergistic Model for Supporting and Testing Students. Proceedings of 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), 407–412. <https://doi.org/10.1109/COMPSAC.2017.211>
- Duval, E., Sharples, M., & Sutherland, R. (2017). Technology Enhanced Learning. Research Themes. Springer International Publishing AG 2017. <https://doi.org/10.1007/978-3-319-02600-8>
- Norman, S. (2016) 5 Advantages of Online Learning: Education Without Leaving Home. eLearning Industry [online]. Retrieved from: <https://goo.gl/jzzyUV>

## Authors



**Marina Marchisio** - [marina.marchisio@unito.it](mailto:marina.marchisio@unito.it)

Marina Marchisio is Professor of Mathematics at the University of Turin, Department of Molecular Biotechnology and Health Sciences. Her research domain is Digital Education, in particular teaching and learning STEM disciplines with new technologies. She is the scientist-in-charge of the Italian Ministry of Education's Project Problem Posing and Solving and the coordinator of several Projects of the University of Turin for digital university guidance, e-learning, school and academic success.

**Sergio Rabellino** - [sergio.rabellino@unito.it](mailto:sergio.rabellino@unito.it)

Sergio Rabellino is the head of the ICT Office of the Department of Computer Science, University of Turin. He co-operates with the research groups in Security, Eidomatics, High Performance Computing, Artificial Intelligence and E-learning. He is a Moodle Developer and hardware/software architect of elearning platforms. Technical Head of Start@Unito, Orient@amente, iLearn and PPS moodle platforms, he has more than 40 publications about elearning tools and methods.





**Matteo Sacchet** - [matteo.sacchet@unito.it](mailto:matteo.sacchet@unito.it)

Besides his studies in Mathematics, he always showed interest in Technology Enhanced Learning, with participations in important projects, both at the local level, at the national level and at the international level. In 2017, for two years, he managed the start@unito project of the University of Turin, training university staff regarding different aspects of creating freely accessible online courses. He is now a Postdoctoral researcher in collaboration with international partners.

**Daniela Salusso** - [daniela.salusso@unito.it](mailto:daniela.salusso@unito.it)

Daniela Salusso holds a PhD in English Studies from the University of Turin. She teaches English Language and Translation at the Department of Foreign Languages and Cultures of the same university as an adjunct professor. In the year 2018-2019, she collaborated in quality of coordinator of the start@unito project. Besides her research in applied linguistics and translation, she is also interested in instructional design, open education, teacher training, and online teaching and learning.



# Scuole connesse e didattica digitale. La scuola del futuro e il progetto Riconessioni

Giovanni Luca Spoto, Flavio Renga, Marcello Enea Newman

Fondazione per la Scuola della Compagnia di San Paolo

**Abstract.** Il seguente articolo presenta una fotografia dello stato dell'arte della connettività delle scuole italiane del primo ciclo. Successivamente, illustra le caratteristiche principali del progetto Riconessioni, un programma di innovazione infrastrutturale e pedagogica che dimostra le opportunità offerte dalla didattica digitale. Il progetto della Compagnia di San Paolo, realizzato dalla Fondazione per la Scuola, ha l'obiettivo di stimolare l'innovazione delle scuole del primo ciclo fornendo accesso alla rete e formazione docenti all'avanguardia

**Keywords.** Innovazione, Connettività, Istruzione.

## Introduzione

L'Italia si è connessa a Internet per la prima volta nel lontano 30 aprile 1986, abbracciando pionieristicamente la tecnologia che forse più di ogni altra ha definito il nostro ingresso nel 21esimo secolo. Questa prima connessione, con una linea di soli 28kbps costituì un ponte tra il Centro universitario per il calcolo elettronico del CNR di Pisa – CNUCE e la stazione di Roaring Creek in Pennsylvania. Pochissimi paesi europei si erano già connessi: Norvegia, Regno Unito e l'allora Germania dell'ovest.

Nonostante questo primato, l'Italia oggi è un paese con basse prestazioni nell'adozione delle ICT rispetto alle altre nazioni europee. Infatti l'Italia è 25esima in Europa per il livello di digitalizzazione del paese. A rivelarlo è l'edizione 2018 del Digital Economy and Society Index (DESI) (EC, 2019), ovvero l'indice complessivo attraverso cui la Commissione Europea misura, attraverso cinque aree politiche principali (connettività, capitale umano, uso di internet, integrazione delle tecnologie digitali e servizi pubblici digitali), il livello di attuazione dell'Agenda Digitale da parte dei singoli Stati membri.

Gli indicatori rivelano anche che in Italia la domanda di banda larga veloce e ultraveloce è in rapido aumento ma non abbastanza da far fronte a esigenze in costante crescita. Questa carenza, ovviamente, riguarda anche il mondo della scuola (Battista, 2018).

Da una recente analisi sui dati MIUR/Agi (Tola, 2017), sappiamo infatti che solo il 15% delle scuole italiane ha la fibra, mentre la maggioranza ha solo l'Adsl. Le scuole che hanno la velocità di banda più efficiente, ossia oltre i 30 Megabit, sono appena una scuola su dieci, nonostante oltre il 60% di loro sarebbero potenzialmente coperte da infrastrutture Fttc/Ftth già poste in essere.

La carenza di connettività costituisce per le scuole italiane un problema particolarmente grave, soprattutto considerando quanto questo rischi di frenare quanto fatto negli

ultimi anni per introdurre pratiche di didattica innovativa e digitale nelle scuole di ogni ordine e grado. In particolare, il sistema dei PON, il Piano Nazionale Scuola Digitale, e la stessa Buona Scuola hanno fatto tantissimo per fornire strumenti e competenze mirati a innovare la scuola pubblica italiana. È evidente che senza un'infrastruttura ICT adeguata l'effetto di queste iniziative sia considerevolmente ridotto.

A pagarne le conseguenze sono tutti gli attori che compongono la comunità educante. La mancanza di connettività, infatti, rappresenta un enorme freno alla trasformazione digitale delle segreterie, all'innovazione delle pratiche didattiche e alla creatività e all'accesso ai saperi dei nostri ragazzi. Dall'altro canto, laddove troviamo connettività e docenti formati, le scuole italiane stanno dimostrando una fortissima capacità di innovarsi e di sperimentare.

## 1. Il progetto Riconessioni

In questa direzione si colloca il lavoro del progetto Riconessioni che, secondo un approccio sistemico organizzato su due fronti complementari, punta a: estendere i collegamenti ad alta capacità per le scuole (per garantire l'equità dell'accesso) e formare i docenti (per diffondere il saper-fare). Nello specifico, il progetto triennale della Compagnia di San Paolo, realizzato dalla Fondazione per la Scuola, porterà la fibra ottica a 250 plessi scolastici entro il 2020 e coinvolgerà più del 50% dei docenti delle scuole primarie e secondarie di primo grado di Torino e della sua area metropolitana (circa 350 scuole) in un esteso programma di formazione sulle competenze digitali e la didattica innovativa. Ogni scuola collegata sarà dotata di una connessione fino a 10 Gbps simmetrici, ulteriormente potenziabili in futuro senza bisogno di modificare l'infrastruttura posata oggi.

Ad oggi il programma ha già coinvolto direttamente più di 800 docenti (e indirettamente più di 2000 altri attraverso formazione a cascata) e connesso più di 100 plessi scolastici alla fibra ottica. Questo lavoro ci sta permettendo di sperimentare nuove soluzioni didattiche che vanno dalla produzione e condivisione di filmati in stop motion, all'uso della realtà virtuale nell'educazione alla cittadinanza globale, alla produzione di unità didattiche sulla storia di internet e sulla supply chain della componentistica dei telefoni cellulari.

Tra le innovazioni didattiche sperimentate nel progetto, uno spazio particolare è riservato al tema del Pensiero Computazionale. Questo è l'insieme dei processi mentali che vengono posti in atto nella formulazione di un problema e nelle azioni necessarie per arrivare alla sua soluzione, adottando una opportuna strategia.

È possibile sperimentare e allenarsi al Pensiero Computazionale sia adottando strumenti digitali sia utilizzando modalità di tipo unplugged, ovvero metodi che non richiedono l'utilizzo di dispositivi elettronici. Nel progetto Riconessioni vengono utilizzate entrambe le modalità, ponendo tuttavia una maggiore attenzione all'utilizzo di quegli strumenti digitali che risultano particolarmente efficaci per lo sviluppo di tale competenza.

Ad esempio viene proposto l'utilizzo dell'applicativo Scratch, un ambiente grafico di programmazione sviluppato dal Lifelong Kindergarten del MIT Media Lab, che risulta

particolarmente adatto ad acquisire e sperimentare le nozioni che stanno alla base della programmazione. Questo linguaggio è facilmente utilizzabile anche nella programmazione di dispositivi di robotica educativa (quali ad esempio Arduino, mBot, micro:bit, LEGO Mindstorms EV3), con i quali è possibile fornire all'utente un importante valore aggiunto all'esperienza didattica: il feedback immediato e tangibile. Questa caratteristica facilita l'apprendimento, riducendo ulteriormente lo sforzo di astrazione necessario all'esercizio del pensiero computazionale e rendendolo tangibile e visibile negli effetti sui robot programmati.

Tali strumenti digitali trovano applicazione sia nelle materie scientifiche (matematica, geometria, logica) sia nelle materie più squisitamente umanistiche. È infatti possibile utilizzare Scratch per raccontare storie, per approfondimenti in ambito storico o geografico, per realizzare attività didattiche utili all'apprendimento della lingua straniera o per agevolare la fruizione di contenuti da parte di ragazzi con difficoltà di apprendimento (ad esempio utilizzando la funzionalità di sintesi vocale abbinata alla visualizzazione di testi). Sia gli aspetti tecnologici che quelli legati alla formazione del progetto Riconessioni sono fortemente basati sulla collaborazione e il partenariato con altri attori dell'ecosistema scolastico che condividono simili obiettivi. Nello specifico, il piano di formazione è arricchito da partenariati con le grandi case editrici di scolastica (incluse De Agostini, Giunti, Pearson, FME Education). La connessione, invece, è erogata grazie a tre partenariati principali con GARR, TOP-IX e Open Fiber.

Le scuole coinvolte nel progetto sono veicolate sulla rete GARR, attraverso una rete di trasporto messa a punto da Open Fiber e gestita da TOP-IX. Il traffico IP da e verso le scuole è veicolato attraverso un collegamento diretto a 10Gbps tra il PoP GARR e la sede TOP-IX a Torino. Per ciò che riguarda l'infrastruttura fisica, queste scuole sono connesse tramite fibra ottica dedicata secondo un'architettura di tipo punto-punto grazie a un accordo con Open Fiber, main partner del progetto, permettendo una connettività rapida e sicura.

## 2. Conclusioni

Innovare la scuola significa dare ai cittadini del futuro le competenze di cui avranno bisogno per rendere la società migliore: più equa, più inclusiva e più creativa. In questo la rete gioca un ruolo centrale, rappresentando la principale condizione abilitante per le innovazioni didattiche di cui i nostri ragazzi hanno bisogno. La connettività delle scuole non è quindi circoscrivibile all'ambito infrastrutturale, ma va discussa e presa in considerazione da chiunque sia interessato all'innovazione e al progresso della società in generale.

## Riferimenti bibliografici

Battista, C. Connessione di qualità per le scuole. Bricks, 2018.

EU Commission, Digital Transformation Scoreboard 2018, EU businesses go digital: Opportunities, outcomes and uptake. EC, 2018.

Tola, E. Poco connesse e a bassa velocità: le #scuole digitali sono ancora un miraggio. AGI, 2017.

## Autori



**Giovanni Luca Spoto** - [giovanni.spoto@fondazione scuola.it](mailto:giovanni.spoto@fondazione scuola.it)

Laureato in Ingegneria Informatica, dal 2009 si occupa di ricerca a sostegno delle tecnologie ICT. Appassionato di Hardware, in Riconessioni cerca soluzioni tecniche a supporto dei processi innovativi.

**Flavio Renga** - [flavio.renga@fondazione scuola.it](mailto:flavio.renga@fondazione scuola.it)

Ingegnere elettronico con una grande passione per la divulgazione scientifica e le idee geniali. Si occupa di ricerca e innovazione, ama declinare la tecnologia e la creatività in ogni possibile forma.



**Marcello Enea Newman** - [marcello.newman@fondazione scuola.it](mailto:marcello.newman@fondazione scuola.it)

Marcello si occupa di didattica innovativa, formazione docenti e relazioni internazionali all'interno del progetto Riconessioni. Ha conseguito una laurea triennale in filosofia all'Università degli Studi di Roma La Sapienza e un master in imprenditoria culturale e creativa presso la Goldsmiths University di Londra.



# Ecosistema Digitale E015: condividere l'accesso regolamentato ai dati per costruire relazioni

Emiliano Sergio Verga, Maurilio Zuccalà

Cefriel – Politecnico di Milano

**Abstract.** Una delle principali sfide per sfruttare al meglio il potenziale delle informazioni disponibili nella smart city e costruire servizi innovativi per cittadini e turisti è quella di semplificare e promuovere lo scambio regolamentato dei dati fra soggetti diversi. Se da un punto di vista tecnico esistono standard di interoperabilità maturi e consolidati, il problema principale rimane la costruzione di un framework di processo che favorisca l'incontro fra soggetti interessati a condividere dati e soggetti interessati a utilizzarli con reciproco vantaggio. Nell'articolo si analizzerà il caso di studio dell'Ecosistema Digitale E015.

**Keywords.** Ecosistema Digitale, Interoperabilità, Smart City, API Economy

## Introduzione

Le nuove tecnologie stanno permeando sempre di più la realtà quotidiana, arricchendo l'offerta di servizi innovativi per gli utenti finali; allo stesso tempo, la diffusione estensiva delle tecnologie legate all'Internet of Things sta producendo sempre più grandi quantità di dati in tempo reale. Questi dati possono diventare preziosi "building block" nella costruzione di nuovi servizi innovativi, mostrandoli opportunamente agli utenti finali attraverso app o altre soluzioni digitali, oppure integrandoli con altri dati attraverso algoritmi per estrarre nuove informazioni a valore aggiunto.

Se da un punto di vista tecnico lo scambio dati è oggi facilmente realizzabile grazie alle tecnologie consolidate dell'interoperabilità dell'API Economy, la vera sfida riguarda tutti gli aspetti di regolamentazione, culturali e di business correlati all'opportunità di scambiare i dati. È dunque necessario prevedere un framework che non consideri soltanto gli aspetti tecnologici per promuovere tale scambio dati: è necessario definire un "Ecosistema Digitale" (Bonardi et al. 2016)(Zuccalà e Verga 2018) all'interno del quale i diversi soggetti, sia pubblici che privati, in accordo alle specifiche esigenze come gli attori di un ecosistema fisico, possano liberamente scambiarsi i dati con reciproco vantaggio in modo semplice e sicuro. Si riporta nel seguito il caso studio dell'Ecosistema Digitale E015 realizzato con il coordinamento scientifico di Cefriel e gestito da Regione Lombardia.

## 1. L'iniziativa "Ecosistema Digitale E015"

L'Ecosistema Digitale E015 ([www.e015.regione.lombardia.it](http://www.e015.regione.lombardia.it)) ha l'obiettivo di favorire lo scambio di dati e funzionalità fra soggetti distinti, sia pubblici che privati, in modo semplice e condiviso. I partecipanti a E015, che hanno sottoscritto un contratto gratuito di

adesione, possono svolgere due ruoli non mutuamente esclusivi:

- API E015 Provider: questo tipo di attore può valorizzare il proprio patrimonio digitale condividendo l'accesso regolamentato alle proprie banche dati nella forma di API E015; le API E015 sono API di tecnologia SOAP o REST, arricchite con informazioni sulle regole di utilizzo e, una volta validate da un organo di governo dell'Ecosistema E015 denominato Technical Management Board, sono censite all'interno di un catalogo di API E015 dell'Ecosistema e richiedibili dagli altri partecipanti.
- App E015 Provider: questo tipo di attore può arricchire le soluzioni software per i propri utenti (siti web, app mobile, totem informativi, cruscotti di monitoraggio ecc.) chiedendo, in accordo alle regole di E015, le funzionalità e le informazioni condivise dagli altri partecipanti nella forma di API E015.

Ogni volta che una API E015 viene utilizzata con successo all'interno di una soluzione digitale di un altro partecipante in uno scenario win-win di interesse comune, nasce una "relazione digitale" fra i due soggetti: a settembre 2019 sono più di 400 le relazioni digitali nate all'interno dell'Ecosistema Digitale E015.

Nato in occasione di Expo 2015 per volere di Confindustria, Confcommercio, Camera di Commercio di Milano, Assolombarda e Unione del Commercio, con il coordinamento scientifico di Cefriel ([www.cefriel.com](http://www.cefriel.com)), E015 è oggi gestito da Regione Lombardia e utilizzato come strumento operativo per promuovere il territorio e mettere a sistema informazioni utili per i cittadini.

## 2. Alcuni esempi di utilizzo in diversi settori

L'Ecosistema Digitale E015 ha trovato utilizzi concreti in diversi settori. In ambito trasporti, per esempio, diverse aziende di gestione aeroportuale e di trasporto ferroviario o su gomma condividono i loro dati e consentono di conoscere in tempo reale (tramite colonnine, display, app) se la situazione dei trasporti è regolare o viceversa se ci sono ritardi o problemi di altro tipo. Le stesse informazioni sono raccolte in una vista integrata per il cittadino nel Portale Infomobilità del Comune di Milano ([www.muoversi.milano.it](http://www.muoversi.milano.it)). Inoltre, le informazioni sul trasporto pubblico locale condivise all'interno dell'Ecosistema sono state utilizzate dalla Fondazione Links, insieme con Rete Ferroviaria Italiana e Regione Piemonte, per valutare costi e benefici dell'inserimento di una fermata intermedia

Fig. 1  
Portale  
Infomobilità  
del Comune  
di Milano

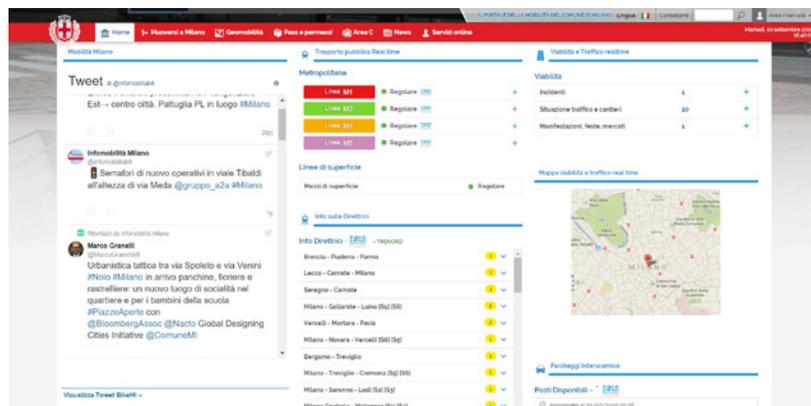


Fig. 2  
Il palinsesto  
dell'Ecosistema  
Digitale del Turismo  
InLombardia  
visualizzato in una  
App E015 in realtà  
aumentata



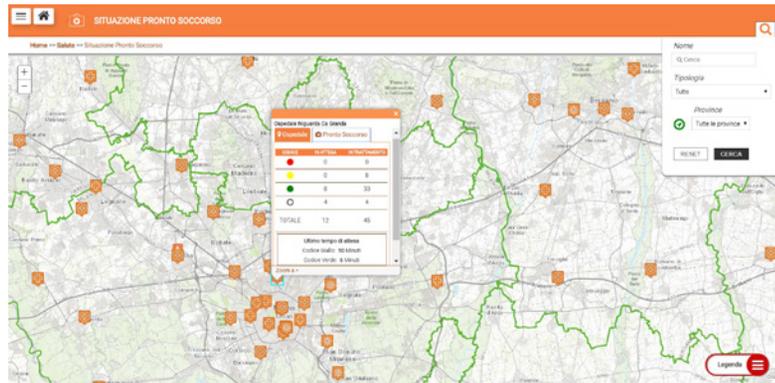
dell'Alta Velocità lungo la tratta Milano-Torino.

E015 è adottato ampiamente nella valorizzazione del patrimonio turistico e culturale del territorio. Explora, la Destination Management Organization di Regione Lombardia e delle Camere di Commercio lombarde, aggrega all'interno dell' Ecosistema Digitale del Turismo InLombardia ([www.inlombardia.it](http://www.inlombardia.it)) le informazioni sugli eventi dei territori, prelevandole da diverse API E015 per ricondividere a sua volta all'interno di E015 un ricco palinsesto integrato che può essere utilizzato da altri soggetti per diverse attività di promozione.

E015 è utilizzato anche in ambito ambientale. I dati condivisi in tempo reale dalle stazioni meteo dell'Agenzia regionale per la protezione dell'ambiente della Lombardia vengono impiegati nel monitoraggio della diga di Pagnona, per predire gli apporti idrologici in arrivo all'invaso. Gli stessi dati vengono utilizzati dalle Regioni del Bacino Padano del Po (Piemonte, Lombardia, Veneto, Emilia-Romagna) nell'ambito dell'“Accordo Aria” per coordinare le azioni di mitigazione dell'inquinamento atmosferico. Le informazioni sull'attivazione o meno di queste limitazioni sono a loro volta ricondivise all'interno dell'Ecosistema Digitale E015: vengono utilizzate, ad esempio, nelle dashboard innovative del progetto SPICA (Zuccalà et al. 2019) destinate agli inquilini di edifici oggetto di interventi di riqualificazione edilizia. Grazie a tale cruscotto, gli inquilini possono conoscere come il loro comportamento ha effetti sui consumi energetici, anche in relazione a eventuali limitazioni sulle temperature massime interne imposte in relazione ai limiti dell'Accordo Aria. Il modello operativo di E015 è anche alla base della gestione della sicurezza: attraverso il “Cruscotto Emergenze”, una dashboard integrata utilizzata dalle Istituzioni all'interno delle sale di controllo, si ha una vista unitaria dei fenomeni del territorio grazie ai dati condivisi da ciascuna istituzione nella forma di API E015.

Infine, l'Ecosistema Digitale E015 trova sempre di più un'utilità nel settore del welfare, ad esempio favorendo lo scambio di dati in tempo reale sullo stato dei pronto soccorso, sulla posizione dei defibrillatori, sugli orari di visita dei medici di famiglia, sulla apertura

Fig. 3  
Lo stato in tempo  
reale del pronto  
soccorso



delle farmacie di turno o sull'offerta di prestazioni sanitarie del territorio.

### 3. Conclusioni

In questo articolo si è illustrato un caso reale di Ecosistema Digitale che permette a soggetti diversi, sia pubblici che privati, di scambiarsi facilmente i dati con reciproco vantaggio e in modo regolamentato. Grazie all'approccio presentato è stato possibile per diversi attori del territorio valorizzare i propri asset digitali e costruire soluzioni innovative per cittadini e turisti. Fra i principali riscontri ottenuti dalla presente esperienza si può sottolineare come la condivisione di API E015 non abbia richiesto particolari investimenti economici, né siano stati necessari interventi onerosi sui sistemi esistenti, promuovendo in questo modo un approccio facilmente scalabile e sostenibile per attori di diverse dimensioni.

Le principali difficoltà riscontrate riguardano la barriera di tipo culturale nel diffondere l'opportunità di partecipazione ad attori di settori differenti: se la fluidità dell'approccio presenta intrinsecamente il vantaggio di superare i confini fra i diversi settori, da un altro punto di vista si è evidenziata la necessità di creare dei casi di successo con un ruolo di dimostratore nei diversi settori, al fine di coinvolgere i soggetti attivi di quell'ambito specifico.

In tal senso il prossimo passo su cui si andrà a lavorare sarà la costruzione di "Ecosistemi Verticali" montati sopra un Ecosistema Digitale abilitante come E015, al fine di guidare e promuovere l'uso dell'approccio E015 all'interno di uno specifico ambito ma senza perdere il vantaggio della trasversalità di E015.

### Riferimenti bibliografici

- Bonardi M., Brioschi M., Fuggetta A., Verga E.S., Zuccalà M. (2016), Fostering Collaboration Through API Economy: The E015 Digital Ecosystem, 3rd International Workshop on Software Engineering Research and Industrial Practice, in conjunction with ICSE 2016 SER&IP'16, ACM, pp. 32–38
- Zuccalà M., Verga E.S. (2018), Governing Interoperability Through Digital Ecosystems: The E015 Experience, IEEE International Conference on Smart Computing (SMART-COMP), pp. 429–433
- Zuccalà M., Mauri M., Cristofaro M., Hugony C., Hugony M.E., Roncelli I. (2019), Actio-

nable data to improve energy efficiency and citizen awareness, 5th Italian conference on ICT for Smart Cities and Communities, I-CiTies 2019

## Autori



**Emiliano Sergio Verga** - [emiliano.verga@cefriel.com](mailto:emiliano.verga@cefriel.com)

Emiliano Sergio Verga è laureato in Ingegneria Energetica presso il Politecnico di Milano e ha conseguito un Master in IT Governance. Lavora in Cefriel come Digital Ecosystem Manager e in particolare si occupa del coordinamento scientifico dell'Ecosistema Digitale E015 di Regione Lombardia. Svolge inoltre il ruolo di docente per diversi master del Politecnico di Milano.

**Maurilio Zuccalà** - [maurilio.zuccala@cefriel.com](mailto:maurilio.zuccala@cefriel.com)

Maurilio Zuccalà è laureato in Ingegneria Informatica presso il Politecnico di Milano. Lavora al Cefriel dal 2001 su attività di ricerca, innovazione e formazione. È stato ed è coinvolto in diversi progetti di ricerca nazionali e internazionali (FP6, FP7, H2020, EIT Digital). I suoi attuali interessi di ricerca comprendono tematiche smart city, API Economy, piattaforme digitali ed ecosistemi a servizi. Ha collaborato alle fasi iniziali della realizzazione dell'Ecosistema Digitale E015.



# Il catalogo che cresce: gli autori della Beic Digital Library in Wikidata

Federica Viazzi

Fondazione BEIC

**Abstract.** Presentazione delle fasi di lavorazione di una delle attività del progetto GLAM/BEIC: l'esportazione in Wikidata dei record relativi agli autori della Biblioteca Digitale BEIC. Vengono descritti i passaggi tecnici necessari al riversamento dei dati nel database. Prima la preparazione dei metadati associati agli autori persona, la loro importazione in uno strumento wiki che permette di confrontare i record presenti in un catalogo con gli elementi esistenti in Wikidata, e di abbinarli nel caso si riferiscano alle stesse entità: Mix'n'match. Successivamente, l'inserimento tramite QuickStatements, un altro tool che permette di intervenire attraverso procedure semi-automatiche, di coppie proprietà-valore secondo la struttura propria a Wikidata. L'ultima parte descrive infine la creazione di nuovi elementi Wikidata, nel caso in cui non vi fossero corrispondenze tra i record del catalogo e il database, sempre tramite lo strumento QuickStatements.

**Keywords.** Metadati, Wikidata, Biblioteche, Digital humanities

## Introduzione

La Fondazione BEIC partecipa dal 2014, a un progetto GLAM-wiki (Galleries, Libraries and Museums) in collaborazione con l'associazione Wikimedia Italia, il capitolo nazionale della Wikimedia Foundation. Nell'ottica di una condivisione aperta dei contenuti e di diffusione della conoscenza libera, ha l'obiettivo di disseminare le risorse della Biblioteca Digitale all'interno di Wikipedia e dei progetti fratelli.

Dall'inizio del GLAM a oggi, le attività hanno spaziato dalla contribuzione alle voci dell'enciclopedia libera, al caricamento di immagini digitalizzate in Wikimedia Commons, alla collaborazione con i progetti di Wikisource. Negli ultimi anni, tuttavia, è cresciuto l'interesse verso Wikidata, tra i progetti fratelli di Wikipedia quello più affine al mondo bibliotecario.

## 1. Wikidata

Wikidata è un database libero, collaborativo, multilingue e secondario che raccoglie dati strutturati per fornire supporto a Wikipedia, a Wikimedia Commons, agli altri progetti del movimento Wikimedia, e a chiunque nel mondo.

Si tratta del progetto che più di tutti si avvicina ai modelli di gestione dei dati bibliografici: esso infatti assume fra le basi concettuali della sua struttura principi che sono fondamentali nella teoria della catalogazione, come il controllo di autorità. L'affinità di Wikidata con le prassi di lavoro bibliotecarie è ulteriormente confermata dal fatto che le entità descritte in Wikidata hanno cominciato precocemente ad arricchirsi degli identificativi

persistenti provenienti da dataset bibliografici autorevoli, come quello della Bibliothèque nationale de France o il CERL Thesaurus.

La decisione di riversare i metadati di autorità della Biblioteca digitale BEIC in Wikidata risale al 2017. I metadati BEIC sottoposti a controllo di autorità sono molti: ovviamente gli autori, ma anche gli editori, i luoghi, i titoli di opere e di espressioni.

## 2. Le fasi di lavorazione

L'insieme di partenza ha incluso i soli autori di tipo persona che ricorressero come intestazioni principali dei record bibliografici, per un totale di circa 5000 nomi; successivamente, si è deciso di includere anche tutti gli autori persona con intestazione secondaria, arrivando a comprendere un set di quasi 15000 nomi.

### 2.1 Preparazione dei dati

Per arrivare al riversamento dei dati di autorità in Wikidata, la fase preliminare necessaria è stata la preparazione dei metadati di partenza. La Biblioteca digitale BEIC utilizza lo standard MARC21 per la creazione dei record bibliografici, pertanto tutte le informazioni relative agli autori di tipo persona sono state rintracciate all'interno dei tag di tipo 100 per le intestazioni principali, e di tipo 700 per le intestazioni secondarie. Ciascuno di questi campi è arricchito con informazioni strutturate che permettono l'identificazione univoca della persona: nome e cognome, date di nascita e morte, eventuali titoli associati, numerali nel caso di cariche politiche (ad es. papi e imperatori).

È importante segnalare che il protocollo catalografico BEIC prevede l'inserimento delle date nel record di autorità in ogni caso, e non soltanto per disambiguare gli omonimi. Questo dato stabile ha avuto grande utilità nel progetto di riversamento in Wikidata, poiché ha permesso un'identificazione più certa e al contempo ha assicurato la presenza di un dato di tipo enciclopedico in tutte le registrazioni.

I metadati associati agli autori persona sono stati esportati dal catalogo, rielaborati e riversati in un file unico, che è stato a sua volta importato in uno strumento wiki specifico: Mix'n'match. Creato da Magnus Manske, uno dei membri della comunità wiki internazionale, Mix'n'match è un tool che permette con relativa facilità di confrontare i record presenti in un catalogo con gli elementi esistenti in Wikidata e abbinarli nel caso si riferiscano alle stesse entità.

### 2.2 I due insiemi

Una volta completata questa fase, Mix'n'match ha presentato due insiemi di risultati: uno che comprende tutti gli elementi abbinati, cioè tutti i casi in cui un nome presente nel catalogo di partenza corrisponde a un elemento già esistente in Wikidata; un altro che include tutti gli unmatched, ovvero i nomi presenti nel catalogo di partenza che non hanno alcun abbinamento con entità di Wikidata.

### 2.3 L'arricchimento con l'entità Biblioteca digitale BEIC degli elementi esistenti

Nel primo caso, gli elementi Wikidata coinvolti avrebbero dovuto solamente essere ar-

ricchiti da una nuova coppia proprietà-valore che evidenziasse la corrispondenza con le singole entità del Catalogo BEIC. Per fare ciò, è stato necessario, innanzitutto, creare un elemento Wikidata che descrivesse la Biblioteca digitale BEIC.

Data la mole di dati, sarebbe stato impensabile procedere modificando uno per volta i singoli elementi da arricchire, pertanto si è deciso di utilizzare un altro strumento che permette di intervenire in Wikidata in maniera semiautomatica: QuickStatements. Tramite questo strumento, partendo da un elenco di elementi Wikidata, è possibile inserire lo stesso tipo di informazione in tutti con un'unica operazione. Nel nostro caso, l'informazione da inserire era la forma controllata dell'autore, con la modalità propria di Wikidata, cioè tramite una coppia proprietà-valore. I dati sono quindi stati impostati secondo una struttura che accoppia la proprietà "Descritto nella fonte" al valore "Biblioteca digitale BEIC", a sua volta arricchito da due riferimenti: la forma del nome e l'URL che lancia una ricerca corrispondente nel Discovery tool della biblioteca digitale.

## 2.4 La creazione di nuovi elementi Wikidata

Il secondo possibile risultato del lavoro in Mix'n'Match include tutti i casi in cui non è stato possibile creare una corrispondenza fra un autore BEIC e un elemento Wikidata. Questa eventualità è risultata essere comunque minoritaria e ha determinato una fase di lavoro differente.

Gli elementi da creare non avrebbero dovuto contenere solo la coppia proprietà-valore relativa alla presenza nella Biblioteca digitale BEIC, ma anche le informazioni usate per descrivere le persone: nome, cognome, date di nascita e morte. Queste informazioni sono state desunte dai metadati bibliografici presenti nel catalogo e successivamente normalizzate in una struttura adatta all'immissione in Wikidata.

Oltre a ciò, si è scelto di fare uno sforzo ulteriore: per ciascun elemento sono state specificate le proprietà 'sesso' e 'occupazione', per le quali è stato necessario incrociare informazioni derivanti da fonti diverse.

## 3. Conclusioni

Il risultato finale di questo lavoro vede l'entità "Biblioteca digitale BEIC" associata tramite la proprietà "Descritto nella fonte" a 14666 elementi Wikidata. Il prossimo passo sarà quello di provare a chiudere il cerchio della condivisione: la prospettiva è infatti di integrare le informazioni sempre più corpose e precise presenti nelle entità di Wikidata all'interno dei record di autorità, a partire dagli identificativi persistenti e dalle forme varianti del nome.

Un'ulteriore prospettiva è data dalla volontà di migliorare la struttura dei record di autorità del Catalogo BEIC, creando per ciascuno di essi un identificativo persistente, che dovrà essere il riferimento stabile al di fuori del catalogo e, ovviamente, negli elementi Wikidata. L'adozione di un identificativo semplificherà significativamente le operazioni su Wikidata, in particolare nelle fasi di aggiornamento.

Una volta concluso il progetto sugli autori di tipo persona, l'obiettivo è quello di estendere la prassi fin qui descritta anche agli altri metadati sottoposti al controllo di autorità.

In particolare, si lavorerà sulle forme normalizzate degli autori di tipo ente, degli editori antichi e moderni – includendo anche quelli presenti nel catalogo dell'Archivio della Produzione Editoriale Lombarda, sempre allestito dalla BEIC – e dei titoli di opere ed espressioni.

Concludendo, l'incontro e talvolta lo scontro tra catalogo e Wikidata hanno generato risultati senza dubbio molto positivi: l'esposizione dei dati di autorità della Biblioteca digitale BEIC al di fuori del contesto del catalogo ha permesso il loro arricchimento e contemporaneamente ha dato visibilità alle risorse descritte.

## Riferimenti bibliografici

GLAM su Meta-Wikimedia <https://meta.wikimedia.org/wiki/GLAM>

Introduzione a Wikidata <https://www.wikidata.org/wiki/Wikidata:Introduction/i>

## Autrice



Federica Viazzi - [federica.viazzi@gmail.com](mailto:federica.viazzi@gmail.com)

Dopo la laurea in Conservazione dei Beni Culturali presso l'Università di Pavia si occupa per molti anni di libro antico approfondendone la conoscenza alla British Library di Londra. Parallelamente coltiva interessi per i metadati e le biblioteche digitali. Attualmente occupa il ruolo di responsabile per i metadati strutturali presso la biblioteca digitale della Fondazione BEIC di Milano e contribuisce allo sviluppo del progetto GLAM in collaborazione con Wikimedia Italia.

# European Open Science Cloud: infrastrutture e servizi a supporto della ricerca all'Università degli Studi di Padova

Lorisa Andreoli, Elena Bianchi, Yuri Carrer, Vera Dean, Antonella Zane,  
Michela Zorzi

Università degli Studi di Padova, Sistema Bibliotecario di Ateneo

**Abstract.** Questo documento presenta l'implementazione da parte del Sistema Bibliotecario dell'Università di Padova di un repository per archiviare, preservare e disseminare i dati di ricerca sviluppati all'interno dell'Ateneo. La relazione descrive le premesse del progetto e i passi che il Sistema Bibliotecario ha adottato per produrre l'infrastruttura tecnologica e corredarla di un adeguato servizio di supporto ai ricercatori.

**Keywords.** Data Repository, Open Data, Open Science, FAIR Data, Archivi aperti

## Introduzione

L'Università di Padova è da tempo attiva nel promuovere l'accesso aperto, dotandosi di archivi aperti per la disseminazione e valorizzazione della propria produzione scientifica. A seguito delle recenti raccomandazioni della Commissione Europea per una gestione e un'apertura dei dati finalizzate al riuso, e tenuto conto dell'avvio dell'iniziativa European Open Science Cloud (EOSC), l'Ateneo patavino ha messo in atto delle azioni per dotarsi di un servizio a supporto dei ricercatori per la gestione e l'utilizzo dei dati della ricerca. L'esito di questo processo ha portato alla messa in produzione di un repository di Ateneo per i dati della ricerca conforme alle raccomandazioni della Commissione e all'organizzazione di servizi di supporto all'Open Science.

## 1. Progetto per lo sviluppo di un Servizio di Research Data Repository istituzionale

### 1.1 Analisi preliminare

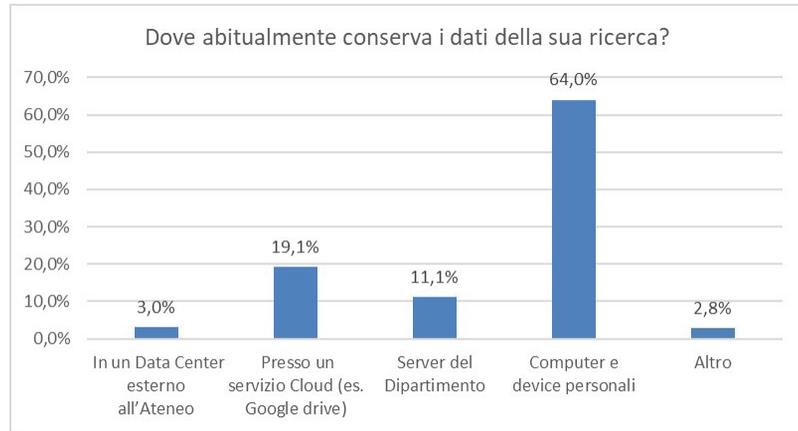
Per conoscere il livello di attuazione delle buone pratiche relative alla gestione dei dati della ricerca, il Sistema Bibliotecario e l'Area Ricerca e Rapporti con le Imprese dell'Ateneo hanno condotto un'indagine conoscitiva tramite interviste e questionari online rivolta alle segreterie di dipartimento e alle comunità di ricerca dell'Ateneo. Tali comunità comprendono ricercatori e docenti che necessitano di un servizio per il deposito e la conservazione dei dati di ricerca alternativo a Zenodo o ad altri repository disciplinari.

I risultati dell'indagine evidenziano che la conservazione dei dati della ricerca è gestita in modo disomogeneo da ricercatori e docenti. Nella maggior parte dei casi i dati sono ar-

chiviati in dispositivi e supporti personali; in minima parte vengono utilizzate infrastrutture messe a disposizione del dipartimento; in alcuni casi ci si avvale di servizi esterni, gratuiti o a pagamento.

Dall'indagine emerge inoltre l'esigenza da parte dei ricercatori di avere a disposizione adeguate infrastrutture tecnologiche e servizi di supporto per la gestione dei dati di ricerca, che vanno dall'intervento di primo livello (help desk) fino a un supporto specifico (Data Management Plan, questioni legali).

Fig. 1  
Conservazione dei dati di ricerca

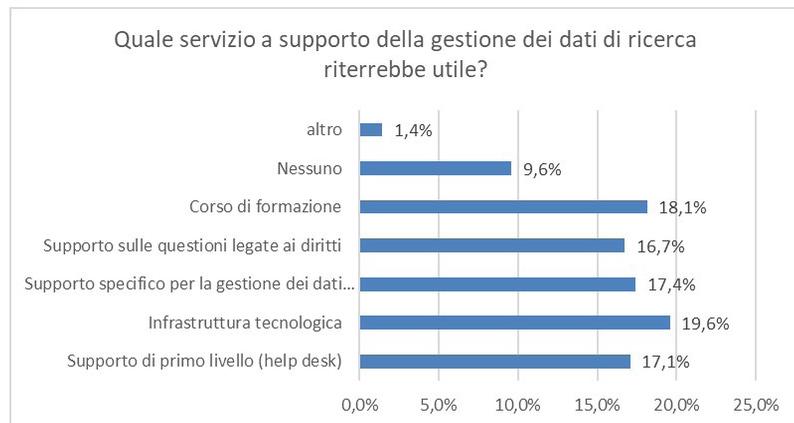


## 1.2 Scelta del software

Il Sistema Bibliotecario di Ateneo ha da anni una dimostrata esperienza nello sviluppo e nella gestione di repository ad Accesso aperto: risale al 2006 la realizzazione di due importanti archivi istituzionali nati per il deposito dei contributi di ricerca e delle tesi di dottorato (Padua@research ) e per il deposito delle tesi di laurea ai fini della disseminazione (Padua@thesis).

Per identificare uno strumento che permetta ai ricercatori di archiviare, conservare e rendere fruibili i dati della ricerca, anche in risposta ai mandati degli enti finanziatori, nel 2017 il SBA ha intrapreso un'analisi dei principali software per repository.

Fig.2  
Servizi richiesti a supporto della ricerca



Dallo studio della letteratura sul tema, si sono individuate le caratteristiche generali e le funzionalità tipiche di un repository istituzionale per i dati della ricerca ad accesso aperto, compresa la possibilità di integrare la nuova piattaforma con l'archivio istituzionale dei prodotti della ricerca (Padua Research Archive).

Tra i software free e open source per repository è stato scelto EPrints già utilizzato in Ateneo per i due archivi sopraccitati. EPrints, sviluppato dall'università di Southampton, è uno dei software per repository più diffuso al mondo.

La gestione del repository (implementazione e sviluppo del software) è curata dal Sistema Bibliotecario, mentre lo storage è ospitato presso il centro dati dell'ASIT (Area Servizi informatici e telematici di Ateneo).

### **1.3 Caratteristiche di Research Data Unipd**

- Accesso aperto;
- Principi FAIR. Dati e metadati rispettano i principi FAIR (Findable, Accessible, Interoperable, Reusable) come richiesto dalla Commissione europea e dai principali enti finanziatori. La Fig. 3 riporta la misurazione della FAIRness del repository;
- Identificatore persistente. Il repository attribuisce in automatico un DOI (Digital Object Identifier) ai dataset depositati, grazie all'adesione del Sistema Bibliotecario di Ateneo al progetto DOI non commerciale della CRUI ;
- Licenze. Ai metadati è associata di default la Licenza Creative Commons 0. Chi deposita può scegliere tra un set di licenze da assegnare ai dataset;
- Soggetti. I dataset vengono classificati utilizzando i settori ERC (European Research Council);
- Integrazione con CRIS (Current Research Information System). Research Data Unipd è tra i primi casi in Italia di integrazione tra un archivio di dati e un archivio di prodotti della ricerca. Gli underlying data depositati in Research Data Unipd sono collegabili con le rispettive pubblicazioni presenti in Padua Research Archive (il CRIS dell'Università di Padova). Allo stesso modo è possibile collegare il dataset al documento pubblicato sul sito dell'editore;
- Deposito. Il deposito dei dataset è consentito agli utenti istituzionali attraverso autenticazione via Single Sign-On di Ateneo;
- Istruzioni per gli autori. La sezione informativa del repository comprende indicazioni generali sulle finalità, sulla policy istituzionale, sui formati dei file e sulle licenze applicabili. Fornisce inoltre istruzioni sulle operazioni preliminari al deposito (verifica dei requisiti previsti dai finanziatori, predisposizione di identificatori personali, URL degli articoli collegati ai dataset, etc.). Nel backend è inoltre disponibile un help contestuale per ogni campo editabile.

### **1.4 Integrazione con infrastrutture e servizi europei relativi ai dati di ricerca**

Per garantire al repository e ai dataset caricati la massima visibilità a livello internazionale, l'archivio è stato indicizzato presso servizi internazionali, quali re3data.org e OpenDOAR. Essendo conforme agli standard richiesti da OpenAIRE , Research Data Unipd è attivo come content provider di questa iniziativa europea.

FINDABLE		assessment	
F1	(meta)data are assigned a globally unique and eternally persistent identifier	✓	F Findable
F2	data are described with rich metadata	✓	
F3	(meta)data are registered or indexed in a searchable resource (able to <a href="#">google data-objects</a> )	✓	
F4	metadata specify the data identifier	✓	
ACCESSIBLE			
A1	(meta)data are retrievable by their identifier using a standardized communications protocol	✓	A Accessible
A1.1	the protocol is open, free, and universally implementable	✓	
A1.2	the protocol allows for an authentication and authorization procedure, where necessary	✓	
A2	metadata are accessible, even when the data are no longer available		
INTEROPERABLE			
I1	(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation	✓	I Interoperable
I2	(meta)data use vocabularies that follow FAIR principles	unclear	
I3	(meta)data include qualified references to other (meta)data	unclear	
RE-USABLE			
R1	(meta)data have a plurality of accurate and relevant attributes	✓	R Reusable
R1.1	(meta)data are released with a clear and accessible data usage license	✓	
R1.2	(meta)data are associated with their provenance	✓	

Fig. 3  
FAIRness di  
Research Data Unipd

### 1.5 La policy per la gestione dei dati della ricerca dell'Università di Padova

Contestualmente al rilascio del repository, il 1 dicembre 2018 il Senato Accademico ha approvato la Policy istituzionale per la gestione dei dati della ricerca, che identifica esplicitamente Research Data Unipd come l'archivio digitale dell'Università di Padova per i dati della ricerca. Il testo della policy si basa sul modello elaborato dal gruppo di lavoro interuniversitario IOSSG (Italian Open Science Support Group).

## 2. La diffusione e la promozione dell'Open Science all'Università di Padova

Attualmente, presso l'Università di Padova, i servizi di supporto alla ricerca per le tematiche relative all'Open Access, agli Open Data e alla normativa sul diritto d'autore vengono garantiti dal Sistema Bibliotecario.

Attraverso il suo sito web fornisce informazioni e approfondimenti sull'Open Access e l'Open Science (Sezione "Per chi pubblica") e il "Servizio Aiuto SBA". La finalità di questo servizio è di offrire al ricercatore un unico punto di accesso a servizi di supporto coordinati tra le varie unità organizzative collegate alla ricerca, come descritto in "A view on the implementation of the European Open Science Cloud".

Inoltre, il Sistema Bibliotecario offre a tutti gli utenti istituzionali (dottorandi, ricercatori, docenti e staff) occasioni di formazione e informazione, in presenza e on-line, sui temi dell'Open Science.

## 3. Conclusioni

Le biblioteche accademiche sono di fronte a un periodo di grandi sfide innovative, quali la gestione e condivisione dei dati prodotti dalla ricerca nell'ambito dell'iniziativa European Open Science Cloud promossa dalla Commissione Europea. In questo contesto, le com-

petenze specialistiche sui dati e metadati, proprie della professionalità del bibliotecario, diventano strategiche per le attività di supporto al “Research data planning and management” essenziali alla ricerca accademica per accedere ai finanziamenti europei.

Lo sviluppo di una infrastruttura per la gestione e archiviazione FAIR dei dati della ricerca e l’offerta di una gamma di servizi a supporto dei ricercatori, il cui accesso è agevolato da un unico punto di contatto, sono i primi necessari passi verso la European Open Science Cloud.

Per l’Università di Padova lo sviluppo della nuova infrastruttura è stata accompagnata e sostenuta da un’azione decisa da parte della governance di Ateneo concretizzata nella adozione di una policy istituzionale specificatamente dedicata ai dati. La formulazione e l’adozione di un documento di policy è infatti il punto di partenza per una piena attuazione della European Open Science Cloud da parte di un’istituzione di ricerca.

## Riferimenti bibliografici

Bianchi E., Budroni, P., Celentano, A., Occioni, M., Toniolo S. (2017), A view on the implementation of the European Open Science Cloud. Conferenza GARR\_2017 Selected Papers. The data way to Science. Venezia, 15-17 novembre 2017, Associazione Consortium GARR, Roma

Solodovnik, Iryna (2015), Repository Istituzionali Open Access e strategie Linked Open Data, Firenze University Press, Firenze [http://www.fupress.com/archivio/pdf/3091\\_9271.pdf](http://www.fupress.com/archivio/pdf/3091_9271.pdf)

<http://bibliotecadigitale.cab.unipd.it/aiuto>

[http://bibliotecadigitale.cab.unipd.it/bd/per\\_chi\\_pubblica/documenti-e-materiali/dichiarazionedimessina2004.pdf](http://bibliotecadigitale.cab.unipd.it/bd/per_chi_pubblica/documenti-e-materiali/dichiarazionedimessina2004.pdf)

[http://bibliotecadigitale.cab.unipd.it/bd/per\\_chi\\_pubblica/documenti-e-materiali/Policy\\_accesso\\_aperto1.pdf](http://bibliotecadigitale.cab.unipd.it/bd/per_chi_pubblica/documenti-e-materiali/Policy_accesso_aperto1.pdf)

<http://bibliotecadigitale.cab.unipd.it/bd/per-chi-pubblica>

<http://paduaresearch.cab.unipd.it/>

<http://tesi.cab.unipd.it/>

<http://v2.sherpa.ac.uk/opensoar/>

<http://www.unipd.it/download/file/fid/55401>

<https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>

<https://elearning.unipd.it/sba/course/view.php?id=21>

<https://erc.europa.eu/>

<https://sites.google.com/view/iossg/home>

<https://www.coar-repositories.org/files/NGR-Final-Formatted-Report-cc.pdf>

<https://www.crui.it/biblioteche-didattica/progetto-doi.html>

<https://www.go-fair.org/fair-principles/>

<https://www.openaire.eu/>

<https://www.re3data.org/>

<https://www.research.unipd.it/>

<https://www.unipd.it/sites/unipd.it/files/2018/policy%20dati%20ricerca.pdf>

<https://zenodo.org/>

## **Autori**

**Lorisa Andreoli** - [lorisa.andreoli@unipd.it](mailto:lorisa.andreoli@unipd.it)

Lorisa Andreoli lavora presso il Sistema bibliotecario dell'Università degli Studi di Padova. È responsabile del Settore Servizi all'utenza e Comunicazione dell'Ufficio Biblioteca digitale. Si occupa dei servizi della biblioteca digitale, in particolare del coordinamento dei progetti di digitalizzazione e della piattaforma "Phaidra - Collezioni digitali", della formazione e supporto alla predisposizione di mostre virtuali, delle attività di supporto relative ai dati della ricerca.

**Elena Bianchi** - [elena.bianchi@unipd.it](mailto:elena.bianchi@unipd.it)

Elena Bianchi lavora dal 1999 per il Sistema bibliotecario dell'Università degli studi di Padova. Dopo alcuni anni come responsabile della Biblioteca Interdipartimentale di Psicologia F. Metelli, dal 2012 al 2018 gestisce il settore Comunicazione del Sistema Bibliotecario. Dal 2018 è a capo dell'Ufficio Biblioteca digitale di Ateneo. Si occupa in particolare di servizi ed infrastrutture tecnologiche a supporto dei ricercatori per l'Accesso aperto a dati e pubblicazioni. E' membro dell'Italian Open Science Support Group (IOSSG).

**Yuri Carrer** - [yuri.carrer@unipd.it](mailto:yuri.carrer@unipd.it)

Yuri Carrer si laurea in Ingegneria Informatica presso l'Università di Padova; nella sua tesi di laurea si occupa di oggetti digitali, attività che continua dal 2003 presso il Sistema Bibliotecario dell'Università di Padova. Sviluppa gli archivi istituzionali Padua@Thesis/Padua@Research e gestisce come responsabile tecnico le collezioni digitali. Recentemente è occupato di dati della ricerca con la piattaforma Research Data Unipd. Ha creato diversi siti con CMS Plone a supporto delle attività istituzionali.

**Vera Dean** - [vera.dean@unipd.it](mailto:vera.dean@unipd.it)

Vera Dean ha un Master in Biblioteconomia e Scienze dell'Informazione. Dopo aver lavorato presso Infologic s.r.l. e Atlantis (oggi ExLibris Italy), lavora dal 2007 per il Sistema Bibliotecario dell'Università degli studi di Padova, dove si occupa del sistema di automazione gestionale del Sistema Bibliotecario. Dal 2018 di occupa anche del Servizio di Supporto alla Ricerca sui temi dell'Open Science e dell'Open Access all'interno dell'Ufficio Biblioteca digitale di Ateneo.

**Antonella Zane** - [antonella.zane@unipd.it](mailto:antonella.zane@unipd.it)

Antonella Zane ha un PhD in Scienze della Terra e 10 anni di ricerca presso l'Università di Padova. Dal 1998 lavora per il Sistema bibliotecario di Ateneo. Dal 2002 collabora con il Digital Library Group dell'ISTI-CNR, Pisa, ed è stata capoprogetto del team padovano del progetto EU "Linked Heritage" (CIP, 2007-2013). Dal 2011 al 2018 è responsabile del Settore Biblioteca digitale con un focus sui repository per i dati della ricerca. Oggi è il responsabile tecnico della Biblioteca biomedica "A. Vallisneri" e membro dell'Italian Open Science Support Group (IOSSG).

**Michela Zorzi** - [michela.zorzi@unipd.it](mailto:michela.zorzi@unipd.it)

Michela Zorzi lavora dal 1999 per il Sistema bibliotecario dell'Università degli studi di Padova. Dal 1999 al 2008 si occupa del Servizio di Prestito Interbibliotecario di Ateneo e successivamente lavora presso la Biblioteca di Scienze Statistiche e le altre biblioteche del Polo di Scienze Sociali. Dal 2018 di occupa del Servizio di Supporto alla Ricerca sui temi dell'Open Science e dell'Open Access all'interno dell'Ufficio Biblioteca digitale di Ateneo.

# Blockchain: aspetti giuridici tra dati, numeri e algoritmi

Fernanda Faini

Università Telematica Internazionale Uninettuno

**Abstract.** Nell'affascinante rapporto che lega tecnologia e diritto, il contributo intende affrontare il fenomeno della blockchain sotto la lente giuridica. A tal fine l'analisi, dopo aver esaminato le caratteristiche che connotano tale tecnologia, tratta gli aspetti giuridici e il contesto normativo europeo e nazionale di riferimento. Il contributo si concentra su alcuni profili particolarmente complessi di interazione della tecnologia blockchain con principi e norme esistenti a tutela dei diritti e dei valori dei nostri ordinamenti, in specifico la disciplina in materia di data protection.

**Keywords.** diritto; blockchain; distributed ledger technology; smart contract; data protection

## Introduzione

Il contributo intende esaminare il rapporto che lega blockchain e diritto, mettendo in luce il quadro normativo di riferimento ed analizzando alcuni profili particolarmente significativi di tale tecnologia, che sollevano problematiche nell'interazione con normative già esistenti a tutela dei diritti, in specifico la disciplina in materia di data protection. Al fine di esaminare la relazione tra scienza giuridica e tecnologia blockchain è necessaria una premessa: in tal caso, come negli altri in cui si confronta con la tecnologia, il diritto è chiamato a dialogare con altri insiemi di regole, in particolare la *lex informatica* o *digitalis*, ossia le regole informatiche; il codice giuridico deve interagire con quello algoritmico per poter essere efficace.

Le regole informatiche, infatti, condizionano i comportamenti umani, dal momento che abilitano azioni e interazioni, collegano effetti, determinano quali informazioni fornire all'utente: la *lex informatica* è capace di condizionare ogni altra forma di regolazione, compresa quella giuridica. Di conseguenza il diritto è chiamato a intervenire sulla tecnologia: ciò che è giuridicamente legittimo deve essere un sottoinsieme del tecnologicamente possibile, al fine di rispettare i principi dell'ordinamento e tutelare i diritti. A tal fine il diritto può decidere di avvalersi della stessa tecnologia, prevedendo misure tecniche capaci di disabilitare azioni illecite.

Pertanto, nel complesso rapporto tra blockchain e diritto, il giurista deve muoversi come un abile equilibrista: la tecnologia non deve prevalere sul diritto, ma il diritto non deve limitare le potenzialità della tecnologia, non solo perché in tal modo non svolgerebbe in modo adeguato la funzione di regolazione, ma anche perché rischierebbe di votarsi all'inefficacia e al mancato rispetto (Finocchiaro 2012; Sartor 2010).

## **1. Tecnologia e diritto**

### **1.1 Aspetti tecnologici**

Alla luce di quanto espresso in premessa, l'analisi giuridica della blockchain deve necessariamente muovere dall'oggetto della regolazione, da ciò che si vuole regolare e, pertanto, dalle caratteristiche della tecnologia blockchain. L'osservazione della tecnologia è necessaria perché proprio nelle sue caratteristiche, come sarà più avanti esaminato, emergono alcune criticità per il diritto.

Seppur con una inevitabile dose di semplificazione, le tecnologie basate su registri distribuiti (distributed ledger technologies o DLT) e la blockchain, species del genus DLT che impiega una catena di blocchi, si caratterizzano principalmente per i seguenti aspetti: disintermediazione, decentralizzazione e distribuzione; immodificabilità, inalterabilità e persistenza dei dati; meccanismo peer-to-peer alternativo di fiducia, consenso e incentivazione; trasparenza, tracciabilità e sicurezza; funzioni di hash, crittografia asimmetrica e validazione temporale (Giuliano 2018).

Tali caratteristiche si attecchiano in maniera diversa nelle diverse tipologie di blockchain, a seconda che si tratti di blockchain permissionless, che si distinguono per essere aperte e liberamente accessibili a chiunque senza autorizzazioni (es. Bitcoin), oppure blockchain permissioned, che hanno la caratteristica di essere chiuse e non accessibili pubblicamente, dal momento che le autorizzazioni sono gestite da un'autorità centrale, oppure, ancora, blockchain ibride, dette anche consorzi, che sono parzialmente decentrate, dal momento che esiste un controllo sul meccanismo di consenso da parte di alcuni nodi preselezionati, che hanno maggiore influenza degli altri (Gambino et al. 2019; Giuliano 2018; Sarzana et al. 2018).

### **1.2 Profili normativi**

Passando dagli aspetti tecnologici ai profili giuridici, le caratteristiche e gli obiettivi che la blockchain permette di perseguire emergono negli atti di riferimento europei e nazionali. A livello sovranazionale, rileva la risoluzione del Parlamento europeo del 3 ottobre 2018 sulle «tecnologie di registro distribuito e blockchain: creare fiducia attraverso la disintermediazione». Secondo la risoluzione, tali tecnologie sono in grado di migliorare l'efficienza dei costi delle transazioni, oltre ad aumentarne la trasparenza; tale paradigma informatico può democratizzare i dati e rafforzare la fiducia, fornendo un percorso sicuro ed efficace per l'esecuzione delle transazioni. L'Unione europea però chiarisce esplicitamente che i rischi e i problemi di tale tecnologia non sono ancora completamente noti.

A livello europeo rileva, altresì, l'istituzione da parte della Commissione europea del EU Blockchain Observatory and Forum nel febbraio 2018 e dell'European Blockchain Partnership nell'aprile 2018.

A livello nazionale, l'art. 8-ter del d.l. 135/2018, convertito dalla legge 12/2019, si occupa di definire le tecnologie basate su registri distribuiti e lo smart contract.

Le «tecnologie basate su registri distribuiti» (DLT), tra le quali rientra la blockchain, sono definite dalla norma, che riprende le caratteristiche tecnologiche sopra esaminate, come «le tecnologie e i protocolli informatici che usano un registro condiviso, distribuito, repli-

cabile, accessibile simultaneamente, architetturealmente decentralizzato su basi crittografiche, tali da consentire la registrazione, la convalida, l'aggiornamento e l'archiviazione di dati sia in chiaro che ulteriormente protetti da crittografia verificabili da ciascun partecipante, non alterabili e non modificabili».

La disposizione chiarisce che la memorizzazione di un documento informatico attraverso l'uso di DLT produce gli effetti giuridici della validazione temporale elettronica di cui all'art. 41 del regolamento UE n. 910/2014. Ai fini della produzione di tali effetti le DLT devono possedere gli standard tecnici individuati dall'Agenzia per l'Italia digitale (AgID). Il secondo comma dell'art. 8-ter definisce lo "smart contract", significativa applicazione della blockchain, come «un programma per elaboratore che opera su tecnologie basate su registri distribuiti e la cui esecuzione vincola automaticamente due o più parti sulla base di effetti predefiniti dalle stesse».

Anche in tal caso, come per le DLT, la norma collega gli effetti giuridici, specificando che lo smart contract soddisfa il requisito della forma scritta previa identificazione informatica delle parti interessate, attraverso un processo avente i requisiti fissati dall'AgID con linee guida (Bomprezzi 2019).

Nelle caratteristiche che normativamente definiscono lo smart contract si scorge un'evoluzione del contratto automatico (es. il distributore automatico di bevande): in tal caso si aggiunge la piena disintermediazione umana e la qualificazione normativa oscilla tra software e documento informatico. Il fatto che lo smart contract configuri un documento informatico comporta la necessità del rispetto delle norme europee e nazionali di riferimento, ossia in particolare il regolamento eIDAS n. 910/2014, il codice civile, il d.lgs. 82/2005 (Codice dell'Amministrazione Digitale) e le relative regole tecniche.

A livello strategico, in Italia è stato nominato dal Ministero dello Sviluppo Economico un gruppo di esperti, formato da 30 profili multidisciplinari, con l'obiettivo di elaborare una strategia nazionale in materia di tecnologie basate su registri distribuiti e blockchain.

## 2. Blockchain e data protection

Per quanto riguarda l'impiego della blockchain, uno dei profili di maggiore criticità è costituito dal rispetto della disciplina in materia di data protection, di cui al regolamento UE 2016/679 e alla normativa italiana (d.lgs. 196/2003 come modificato dal d.lgs. 101/2018). In premessa è opportuno rilevare che nella blockchain la funzione di hash può essere fatta rientrare in un'operazione di pseudonimizzazione, di cui all'art. 4, par. 1, n. 5, del regolamento UE 2016/679, che come tale comporta l'applicazione della normativa in materia di protezione dei dati personali.

Di conseguenza, in caso di utilizzo della blockchain, laddove siano presenti dati personali, è necessario il rispetto dei principi applicabili al trattamento previsti dalla normativa, in specifico dall'art. 5 del regolamento UE 2016/679, tra i quali rilevano ai fini di tale analisi la minimizzazione dei dati (i dati devono essere adeguati, pertinenti e limitati a quanto necessario rispetto alle finalità per le quali sono trattati) e la limitazione della conservazione (i dati devono essere conservati in una forma che consenta l'identificazione degli interessati per un arco di tempo non superiore al conseguimento delle finalità per le

quali sono trattati).

In considerazione delle esaminate caratteristiche tecniche, la blockchain configge con il rispetto di tali principi: sotto il profilo della minimizzazione dei dati, la difficoltà sta nel fatto che la blockchain per il suo funzionamento replica i dati nei vari nodi e, sotto il principio della conservazione, come precisato anche nella norma italiana, i dati nella blockchain sono «non alterabili e non modificabili». Tali aspetti costituiscono indubbiamente punti di forza di tale tecnologia, ma rischiano di trasformarsi in aspetti di debolezza capaci di creare complesse problematiche al cospetto dei principi in materia di data protection (Gambino et al. 2019).

La normativa in materia di protezione dei dati personali individua alcuni soggetti di riferimento della disciplina: in particolare, accanto all'interessato a cui i dati personali si riferiscono, il titolare, ossia la persona fisica o giuridica, l'autorità pubblica, il servizio o altro organismo, che, singolarmente o insieme ad altri titolari, determina le finalità e i mezzi del trattamento, e il responsabile, eventualmente preposto dal titolare (art. 4, reg. UE 2016/679). L'individuazione di tali figure è fondamentale nella disciplina, dal momento che a tali soggetti si applicano obblighi necessari a garantire la tutela della persona e dei suoi dati, come il principio di responsabilizzazione, secondo cui il titolare è competente per il rispetto dei principi e deve essere in grado di provarlo (art. 5, par. 2, reg. UE 2016/679).

In tal caso rileva la tipologia di blockchain, dal momento che nelle permissioned è individuabile il titolare e nel caso dei consorzi si può fare leva sulla norma relativa alla contitolarità (art. 26, reg. UE 2016/679), ma nelle permissionless diventa complesso individuare tali figure: tutti i nodi devono essere considerati contitolari? Non è presente nessun titolare? In tal caso come trova applicazione la disciplina? (Gambino et al. 2019; Giuliano 2018).

Insieme al necessario rispetto dei principi e all'individuazione delle figure di riferimento, è necessario garantire i diritti dell'interessato, che spaziano nel regolamento europeo 2016/679 dall'accesso (art. 15) alla rettifica (art. 16), dalla cancellazione (art. 17) alla limitazione di trattamento (art. 18), dall'opposizione (art. 21) al non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato (art. 22). In considerazione delle richiamate caratteristiche di immutabilità, inalterabilità e persistenza dei dati, nella tecnologia blockchain diventano difficili la modifica, la rettifica, l'opposizione e la cancellazione dei dati stessi e, di conseguenza, il rispetto di tali diritti dell'interessato (Paladino 2019).

### 3. Conclusioni

In considerazione degli ambiti presi in considerazione emergono alcune complesse criticità nel rapporto tra blockchain e diritto. Per quanto attiene all'aspetto documentale e agli smart contracts, risulta necessario un maggiore coordinamento tra la norma italiana dedicata alla blockchain e il quadro di regolazione europea e nazionale in materia di documenti informatici. Sotto tale profilo, particolare rilevanza assumono le linee guida e gli standard tecnici cui è chiamata l'AgID per dare contenuto ai principi posti dalla norma.

Sotto il profilo della data protection, al fine di superare le problematiche esaminate, una direzione è individuabile nell'approccio sistematico, preventivo, proattivo e tecnico, previ-

sto dal regolamento UE 2016/679, che si traduce nell'incorporazione dei principi e delle norme nella tecnica, facendo assolvere al diritto la funzione preventiva che gli spetta: la regolazione giuridica può servirsi della tecnologia, adattandola al fine di garantire il suo rispetto.

La blockchain mostra sicuramente sfide inedite per il diritto, che è necessario affrontare per far esplodere il valore che tale tecnologia è in grado di esprimere. I nodi aperti e la necessità di rispettare i principi e i valori giuridici di riferimento evidenziano l'importanza della capacità umana di orientare la tecnologia: qualunque evoluzione tecnologica da sola non è sufficiente se non è accompagnata anche dallo strumento del diritto, che da sempre costituisce il mezzo con cui l'uomo regola la società.

### Riferimenti bibliografici

Bomprezzi C. (2019), Commento in materia di Blockchain e Smart contract alla luce del nuovo Decreto Semplificazioni, in *Diritto, mercato, tecnologia*, pp. 1-7.

Finocchiaro G. (2012), Riflessioni su diritto e tecnica, in *Il diritto dell'informazione e dell'informatica* (4-5), pp. 831-840.

Gambino A.M., Bomprezzi C. (2019), Blockchain e protezione dei dati personali, in *Il diritto dell'informazione e dell'informatica* (3), p. 619 ss.

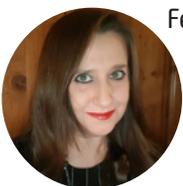
Giuliano M. (2018), La blockchain e gli smart contracts nell'innovazione del diritto nel terzo millennio, in *Il diritto dell'informazione e dell'informatica* (6), p. 989 ss.

Palladino A. (2019), L'equilibrio perduto della blockchain tra platform revolution e GDPR compliance, in *Rivista di diritto dei media* (2), pp. 144-158.

Sartor G. (2010), *L'informatica giuridica e le tecnologie dell'informazione. Corso di informatica giuridica*, II ed., Giappichelli, Torino.

Sarzana di S. Ippolito F., Nicotra M. (2018), *Diritto della blockchain, intelligenza artificiale e IoT*, Wolters Kluwer - Ipsoa, Milano.

### Autrice



**Fernanda Faini** - [fernandafaini@gmail.com](mailto:fernandafaini@gmail.com)

Docente di "Diritto e nuove tecnologie" presso l'Università Telematica Internazionale Uninettuno. Responsabile dell'assistenza giuridica in materia di amministrazione digitale, innovazione tecnologica e informatica giuridica presso la Regione Toscana. PhD in diritto e nuove tecnologie presso l'Università di Bologna, collabora nell'insegnamento di Informatica giuridica presso l'Università degli Studi di Firenze, dove è membro dello Steering Committee dell'Unità di ricerca "Babel – Blockchains and Artificial Intelligence for Business, Economics and Law". Membro del Gruppo di esperti in materia di tecnologie basate su registri distribuiti e blockchain del Ministero dello Sviluppo Economico.

# Un authority file per i luoghi e le istituzioni: l'Atlante Storico Istituzionale dell'Italia unita

Stella Di Fazio

ICAR - Istituto Centrale per gli Archivi

**Abstract.** Il contributo vuole evidenziare il processo di recupero e valorizzazione di dati pubblici e autorevoli rinvenibili in fonti diverse per età e formato alla base della realizzazione dell'Atlante Storico Istituzionale dell'Italia unita (ASI), strumento di valenza generale che si è posto l'obiettivo di documentare le evoluzioni nel tempo delle circoscrizioni territoriali dell'Italia postunitaria, riconducendo progressivamente questo patrimonio informativo eterogeneo ad un modello ontologico unitario, facilmente riusabile in qualunque contesto grazie alla sua pubblicazione in modalità Linked Open Data. L'attuale ontologia ASI, alimentata ad oggi dai dati relativi agli enti territoriali dello stato italiano, sarà gradualmente estesa con l'introduzione nel modello di ulteriori tipologie di circoscrizioni territoriali, nonché del riferimento alle istituzioni che vi hanno operato e alle carte da queste prodotte, in modo da realizzare una dorsale informativa basilare per la descrizione e la contestualizzazione del patrimonio archivistico nazionale.

**Keywords.** Interoperabilità, Linked Open data, Semantic Web, Ontologie geografiche

## 1. Introduzione

Authority files e vocabolari controllati costituiscono architetture informative di supporto alla descrizione e alla contestualizzazione del patrimonio culturale, che agevolano la normalizzazione e il controllo d'autorità dei dati evitando la frammentazione degli interventi e quindi la dispersione o la duplicazione delle attività, ed innalzando al contempo il livello qualitativo dell'informazione prodotta e resa disponibile all'utenza. Per la loro centralità nel processo descrittivo è opportuno che tali strumenti vengano formalizzati in modo tale da risultare facilmente integrabili e/o referenziabili nei diversi contesti informativi: questa è una delle principali motivazioni che hanno indotto l'Istituto centrale per gli archivi (ICAR) a procedere verso la realizzazione progressiva di questo tipo di infrastrutture informative ricorrendo ai formalismi e alle tecnologie del Semantic Web, con l'obiettivo di arrivare gradualmente a costituire una "dorsale informativa" integrata che possa fungere da essenziale punto di riferimento e di snodo per la produzione dell'informazione descrittiva relativa al cultural heritage, e al patrimonio archivistico in primis, come pure per la sua fruizione sul Web.

Alcuni di questi strumenti sono nati all'interno di contesti di fruizione molto ben delimitati, come i vocabolari controllati confluiti nel Tesoro in formato SKOS del Sistema Archivistico Nazionale (SAN) o, sempre in relazione al SAN, il dataset in formato LOD dei soggetti conservatori "d'autorità" della documentazione archivistica in esso censita e descritta. L'Atlante storico istituzionale dell'Italia unita (ASI) si evidenzia invece come uno

strumento di valenza generale, realizzato dall'ICAR con l'obiettivo di rendere disponibile e fruibile in un contesto di uso aperto e interoperabile una grandissima quantità di dati storico-territoriali provenienti da fonti autorevoli di natura pubblica, riconducendola ad un modello ontologico unitario, che consente di rappresentare sia la dimensione diacronica che una visione sincronica delle porzioni di territorio prese in esame.

Censendo le circoscrizioni territoriali dell'Italia postunitaria, a partire dalle circoscrizioni amministrative (base informativa della prima release del maggio 2018) e registrandone i cambiamenti nel tempo, l'ontologia ASI costituisce come esito basilare un repertorio di autorità di nomi di luogo storicamente attestati e georeferenziati, concreto e autorevole supporto per denominare, descrivere e localizzare correttamente le istituzioni collegate a tali luoghi, e le fonti archivistiche da queste prodotte, ma che può allo stesso modo essere utilmente preso a riferimento in qualunque contesto descrittivo, laddove sussista l'esigenza di fornire informazione relativa all'evoluzione storica del territorio. La base di conoscenza in formato RDF OWL è stata alimentata attraverso un lungo processo di studio, analisi e riuso di almeno due diverse tipologie di fonti.

## 2. Il riuso di dati aperti

La principale fonte di alimentazione è costituita dal SISTAT (Sistema Informativo Storico delle Amministrazioni Territoriali), i cui dati sono stati resi disponibili come open data in formato excel dall'ISTAT. Il SISTAT raccoglie e gestisce le informazioni relative agli enti territoriali definiti nell'ordinamento costituzionale italiano (regioni, province, comuni), a partire dal 1861 ai giorni nostri; ne vengono forniti dati quali il nome, la data di istituzione, e quella di ciascuna variazione occorsa nel tempo fino all'eventuale estinzione (ciascuna corredata dai relativi riferimenti normativi), la dipendenza amministrativa da enti gerarchicamente superiori, ecc. Altra fonte di particolare rilevanza, cui si è fatto ricorso al momento solo parzialmente, è costituita da una considerevole quantità di dizionari storici dei comuni, compilati su mandato governativo a seguire l'Unità d'Italia, che a partire dal livello amministrativo comune raccolgono per specifici anni (quindi in tagli sincronici) dati relativi all'assetto istituzionale del territorio (per ogni Comune i riferimenti a Mandamento, Provincia o Prefettura, Circondario amministrativo o Sotto-prefettura, Tribunale di circondario, Corte d'appello, Diocesi).

A partire da tali dati si è avviato un processo di rielaborazione, integrazione, validazione e arricchimento informativo (principalmente conseguito integrando i livelli circondario e mandamento assenti dal SISTAT) che ha portato alla definizione di una ontologia del territorio dello Stato italiano che rispetto ai dati originari presenta alcune significative evoluzioni innovative.

## 3. Ontologia e Linked Open data ASI

Il modello ontologico definito è costruito attorno al concetto di regione spazio temporale, entità astratta attraverso cui vengono descritte le caratteristiche (denominazione, estensione territoriale, dipendenza amministrativa, ecc.) stabilmente attribuibili ad una determinata porzione di territorio (nei termini dell'ontologia unità territoriale) per l'arco

temporale definito come l'intervallo tra un evento che genera un cambiamento e il successivo. Di norma l'evento di cambiamento corrisponde ad un provvedimento normativo che determina la variazione (istituzione, cessione/acquisizione di territorio, mutamento di dipendenza gerarchica, estinzione, ecc.). È il concetto di regione spazio temporale che consente di ottenere, interrogando opportunamente l'ontologia, la rappresentazione esatta della configurazione amministrativa del territorio italiano in un determinato istante nel tempo. Da un punto di vista diacronico, invece, la storia di ogni specifica unità territoriale è costituita dalla successione di tutte le regioni spazio temporali ad essa riferibili.

Attualmente l'ontologia espone i dati relativi a una specifica sottoclasse di unità territoriale, l'unità amministrativa, istanziata da risorse RDF OWL corrispondenti a regioni, province e comuni dello Stato italiano, dall'unità ai giorni nostri; i dati relativi agli altri due livelli amministrativi, circondari e mandamenti, sono stati invece forniti solo per i primi anni successivi alla proclamazione del Regno d'Italia, con l'obiettivo di completarne progressivamente il quadro storico arrivando fino al 1927.

I dati dell'ontologia ASI sono raggiungibili attraverso due distinte piattaforme di pubblicazione e fruizione, una – l'endpoint SPARQL [dati.san.beniculturali.it/sparql](http://dati.san.beniculturali.it/sparql) - destinata a forme di accesso e di riuso dei dati di tipo automatico; l'altra – una applicazione web raggiungibile all'indirizzo <http://dati.san.beniculturali.it/asi/local/> – è costituita da un sistema di ricerca e navigazione a faccette pensato invece per agevolare l'interazione coi dati RDF da parte di utenti umani.

I dati OWL di ASI sono a tutti gli effetti Linked open data, essendo stati inseriti link semantici alle corrispettive risorse delle principali ontologie geografiche disponibili (GeoNames, ISPRA, ISTAT, oltre ovviamente alla base di conoscenza di DBpedia), garantendo così forme di interoperabilità estesa col resto del dominio geografico. Sono stati previsti anche collegamenti semantici tra le risorse RDF rappresentanti le unità territoriali di ASI e le corrispettive risorse presenti nei dataset LOD del SAN (corrispondenti alla sede del soggetto conservatore e del soggetto produttore dei complessi archivistici descritti nel Sistema).

#### **4. Risultati conseguiti e sviluppi futuri**

Già in questa sua prima release l'Atlante Storico Istituzionale dell'Italia unita ha conseguito una serie di vantaggi derivati principalmente dalla rappresentazione in formato RDF dei dati. Un primo risultato di rilievo è rinvenibile nell'aver convogliato e reso immediatamente disponibili in un unico punto di accesso online aperto e di facile utilizzo informazioni complesse arricchite, ridefinite, e potenziate dai processi di interlinking semantico. Un ulteriore vantaggio risiede nella possibilità di riferirsi a ciascuna delle unità territoriali descritte dall'ontologia come risorsa RDF referenziata da un URI, richiamabile all'interno della descrizione di qualunque risorsa culturale (che si gioverà così automaticamente di tutto il potenziale informativo dell'istanza ASI. Infine, con ASI è stato messo a punto uno strumento fortemente caratterizzato dalla interdisciplinarietà: trattando dimensioni informative universali come lo Spazio, il Tempo, ed Eventi di carattere storico-istituzionale collocati su tali direttrici, l'intelaiatura informativa fornita da ASI risulta un fundamenta-

le supporto alla descrizione di qualunque tipologia di risorsa culturale.

L'Atlante Storico Istituzionale dell'Italia unita è stato concepito come un sistema aperto e in espansione: il modello si presta per sua natura ad essere progressivamente esteso con entità, concetti e dati relativi ad altre tipologie di circoscrizioni territoriali (giudiziarie, ecclesiastiche, militari, ecc.), per realizzare una sempre più esaustiva rappresentazione dell'evoluzione storico istituzionale italiana rapportata al territorio. L'ICAR intende dare avvio ad una nuova fase operativa del progetto indirizzata allo studio e all'analisi delle circoscrizioni territoriali di tipo giudiziario e della loro evoluzione nel tempo; in questa fase verrà anche esplicitato formalmente nel modello il legame tra luoghi, istituzioni che in tali luoghi hanno sede, e documentazione archivistica da queste prodotta. Questa estensione dell'ontologia permetterà di porre l'Atlante come base informativa centralizzata per la descrizione e la contestualizzazione del patrimonio archivistico nazionale.

### Riferimenti bibliografici

Bonini, F., Blanco, L., Mori, S., Galluccio, F. (a cura di.) (2016), *Orizzonti di cittadinanza. Per una storia delle circoscrizioni amministrative dell'Italia unita*. Soveria Mannelli, Rubbettino editore, 2016. Cfr. in particolare: Carucci, P., Di Fazio, S., *Dalla Guida generale degli Archivi di Stato italiani all'Atlante storico istituzionale*, pp. 445-460.

Ontologia ASI, <http://dati.san.beniculturali.it/lode/asi/ASI.owl>.

### Autrice

Stella Di Fazio - [stella.difazio@beniculturali.it](mailto:stella.difazio@beniculturali.it)

Archivista, si è dedicata prevalentemente allo studio e all'analisi dei linguaggi formali e degli standard descrittivi archivistici. Si occupa da tempo di rappresentazione della conoscenza attraverso gli standard del Semantic Web, ed ha preso parte a numerosi progetti promossi da amministrazione archivistica, università ed enti di ricerca, che hanno portato alla definizione di ontologie informatiche di ambito archivistico e alla produzione di Linked Open Data. Dal febbraio 2018 è archivista di Stato presso l'ICAR.

