

CONFERENCE
GARR
2021

SOSTENIBILE/ DIGITALE

Dati e tecnologie
per il futuro

SELECTED
PAPERS



CONFERENCE
GARR
2021

SOSTENIBILE/DIGITALE

Dati e tecnologie per il futuro

ONLINE, 7-16 GIUGNO 2021

SELECTED
PAPERS



Conferenza GARR 2021 - Sostenibile/Digitale. Dati e tecnologie per il futuro - Selected Papers
Online, 7-16 giugno 2021

ISBN 978-88-946629-0-0

DOI 10.26314/GARR-Conf21-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

Curatori editoriali: Marta Mieli, Carlo Volpe

Progetto grafico: Carlo Volpe

Impaginazione: Marta Mieli, Carlo Volpe

Prima stampa: Novembre 2021

Numero di copie: 500

Stampa: Tipografia Graffietti Stampati snc

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

Tutti i materiali relativi alla Conferenza GARR 2021 sono disponibili all'indirizzo:

www.garr.it/conf21

Indice

- 6 **Realising social inclusion for persons with disabilities and delivering effective science communication through sustainable ICT solutions: a case study from the University of Pavia**
Silvia Favalli, Flavio Ceravolo, Matilde Ceron, Anna Facchinetti, Giulia Melis,
Vittorio Pampanin, Marco Porta, Beatrice Rabai, Carola Ricci, Francesco Scervini

- 10 **TrashWare e Digital Divide**
Alexander Palummo

- 15 **A brand-new flexible and scalable ENEA solution to handle data in smart city environment**
Giuseppe Santomauro, Marta Chinnici, Giovanni Ponti

- 18 **Ludoteca del Registro .it: i laboratori di cybersecurity**
Giorgia Bassi, Beatrice Lami

- 23 **Cybersecurity: sfide e prospettive per una società più sostenibile**
Angelica Maiorano

- 27 **Carbon Footprint Italy: how blockchain technology supports innovation in tackling climate change**
Daniele Pernigotti

- 31 **An Efficient and Privacy-Aware Method for Revealing Network Covert Channels**
Marco Zuppelli, Luca Caviglione, Corrado Pizzi, Matteo Repetto

- 35 **Riconoscimento vocale di formule**
Matteo Amore, Tiziana Armano, Cristian Bernareggi, Anna Capietto, Sandro Coriasco,
Roberta Crespan, Mattia Ducci, Maria Luisa Gabrielli, Maria Francesca Guadalupi,
Alessandro Mazzei, Antonio Mazzei, Adriano Sofia, Francesco Tarasconi

- 41 **CHNet cloud: an EOSC-based cloud for physical technologies applied to cultural heritages**
Alessandro Bombini, Lisa Castelli, Luca dell'Agnello, Achille Felicetti, Francesco Giacomini,
Franco Niccolucci, Francesco Taccetti

- 47 **The technology transfer in Healthcare 4.0: the use of Machine Learning solutions for neurodegenerative diseases**
Annamaria Demarinis Loiotile, Loredana Bellantuono, Francesco De Nicolò, Alfonso Monaco, Sabina Tangaro, Nicola Amoroso, Roberto Bellotti

- 53 **Didattica Innovativa per l'Applicazione dell'Intelligenza Artificiale in contesti Industria 4.0**
Daniele Mazzei, Riccardo Amadio, Daniele Atzeni

- 58 **L'intelligenza artificiale contro la violenza di genere**
Claudia Villante, Gabriele Paone, Maria Giuseppina Muratore

- 63 **Infrastrutture digitali per i dati della ricerca: adozione e personalizzazione del Research Data Management System all'Università Milano-Bicocca**
Bonaria Biancu, Alessandro Andretto, Paolo Brambilla
- 67 **Il percorso di adozione del paradigma Open Science all'INGV**
Locati M., Mele F.M., Montalto P., Romano V., Lauciani V., Vallone R., Cacciaguerra S.
- 72 **EOSC and Research Infrastructure: Legal Challenges of HPC**
Ludovica Paseri
- 77 **"Data steward" per i dati FAIR**
Valentina Pasquale, Emma Lazzeri, Elena Giglia
- 81 **Integrazione dei Principi FAIR nel processo di ricerca biomedica: la creazione del Registro della Regione Sardegna**
Alessandro Sulis, Vittorio Meloni, Cecilia Mascia, Franco Cappai, Caterina G. Carboni, Ernesto D'Aloja, Giorgio Fotia, Giuseppe Serra, Giovanni Sotgiu, Paolo Uva, Francesca Frexia, Gianluigi Zanetti
- 88 **EOSC-Pillar: dall'analisi giuridica alla definizione di policy per la Scienza Aperta**
Sara Di Giorgio, Nadina Foggetti, Caterina Sganga
- 93 **Didattica digitale ed inclusione: il progetto Suoniamo**
Maria Claudia Buzzi, Marina Buzzi, Marco Maugeri, Gabriella Paolini, Alessandra Sbragia, Caterina Senette, Amaury Trujillo
- 99 **Quando la Scuola cambia, un modello sostenibile**
Angelo Canio D'Alessio, Pasquale Boezio
- 104 **Didattica Online: modello di formazione per docenti di ogni ordine e grado**
Cecilia Fissore, Francesco Floris, Marina Marchisio, Matteo Sacchet, Sergio Rabellino
- 109 **Una piattaforma scalabile per esami sicuri da remoto**
Damiano Perri, Osvaldo Gervasi, Marco Simonetti
- 114 **Reti di apprendenti, reti di apprendimento. Un uso ecologico della social network analysis per la progettazione di LMOOC**
Alessandro Puglisi
- 118 **Riconessioni: competenze digitali e nuove infrastrutture per abilitare l'innovazione dal basso del paradigma educativo e realizzare pienamente l'autonomia scolastica**
Giovanni Luca Spoto, Marcello Enea Newman, Veronica Ruberti, Chiara Ciociola



Comitato di programma

Dave Anilkumar - INFINITE AREA
Claudia Battista - GARR
Andrea Cavalli - IIT
Marta Chinnici - ENEA
Moreno Curti - Policlinico San Matteo Pavia Fondazione IRCCS
Marcello Donatelli - CREA
Stefano Epifani - Sapienza Università di Roma
Speranza Falciano - INFN
Paolo Favali - EMSO
Cesare Furlanello - OROBIX LIFE
Fulvio Galeazzi - GARR
Elena Giglia - Università di Torino
Paola Inverardi - Università dell'Aquila
Giuseppe Lo Re - Università di Palermo
Marco Marletta - GARR
Marta Mieli - GARR
Luisa Minghetti - ISS
Gabriella Paolini - GARR
Federico Ruggieri - GARR
Riccardo Smareglia - INAF
Sabrina Tomassini - GARR
Georg Umgiesser - CNR ISMAR
Davide Vagheti - GARR
Simona Venuti - GARR
Michele Vianello - Esperto smart cities e processi di digitalizzazione
Carlo Volpe - GARR

Tutte le presentazioni e maggiori informazioni
sono disponibili sul sito dell'evento:
www.garr.it/conf21

Realising social inclusion for persons with disabilities and delivering effective science communication through sustainable ICT solutions: a case study from the University of Pavia

Silvia Favalli, Flavio Ceravolo, Matilde Ceron, Anna Facchinetti, Giulia Melis, Vittorio Pampanin, Marco Porta, Beatrice Rabai, Carola Ricci, Francesco Scervini

University of Pavia

Abstract. This paper introduces the research project RISID («Realizing the right to social inclusion for persons with disabilities through new tools of smart communication and sharing knowledge: from international to local effectiveness»), which embraces a multidisciplinary perspective involving experts in law, economics, sociology and engineering. On the basis of the theoretical analysis carried out, a digital tool named "SmartInclusion@unipv" (SI@) will be developed to implement the accessibility of the courtyards of the Palazzo Centrale of the University of Pavia and to effectively share scientific knowledge. The App SI@ is presented as a case study of sustainable ICT

Keywords. Sustainable ICT; Disability; Human rights; SDGs; Social inclusion; Science communication; Accessibility at University

Introduction

This paper introduces the background, objectives and implementation of the project RISID («Realizing the right to social inclusion for persons with disabilities through new tools of smart communication and sharing knowledge: from international to local effectiveness»), financed by Fondazione Cariplo and developed by the University of Pavia in cooperation with the disability NGO Fondazione ASPHI onlus, acting as a consultant.

The research is conducted adopting a multidisciplinary perspective and combining different research methodologies both qualitative (legal and sociological analysis) and quantitative (experimental statistical analysis and utilization of surveys). On the basis of the theoretical analysis carried out, a sustainable ICT (the smartphone App "SmartInclusion@Unipv" or SI@) working within the courtyards of the University of Pavia will also be improved as a final outcome

1. Sustainable ICT solutions: fostering social inclusion and spreading scientific knowledge

The widespread use of information and communications technologies (ICTs), including digital technologies, and global interconnectedness worldwide have the potential to brid-

ge the digital divide and to develop knowledge and inclusive societies (Ricci 2019). New technologies play a pivotal role in the development of the ambitious United Nations 2030 Agenda for Sustainable Development, with particular reference to the realization of goal 3 (ensure healthy lives and promote well-being for all at all ages), goal 4 (ensure inclusive and equitable quality education and promote lifelong learning opportunities for all), goal 9 (build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation), goal 10 (reduce inequality within and among countries) and goal 11 (make cities and human settlements inclusive, safe, resilient and sustainable). However, the ongoing process of digitalisation, which is a key part of the fourth industrial revolution, is not inherently sustainable. This phenomenon can deliver on global sustainability goals only if underpinned by a so-called digital sustainability approach, thus empowering the role of ICTs in ensuring sustainability.

Moreover, digital technologies have acquired a primary role in individual and community life (Foley and Ferri 2012; Lazar 2017). They both represent an unprecedented opportunity to actively participate and be fully included in society for vulnerable groups at risk of social exclusion, such as persons with disabilities, and a crucial means to provide information to all citizens (Ellis and Kent 2016). The use of digital technology to create augmented reality and to encourage a storytelling process of the scientific knowledge also embodies for the scientific community a new way to interact and communicate with citizens (Wilsdon and Willis 2004; Wynne 2006).

2. The project RISID in a nutshell

The project RISID purports to successfully develop new patterns for promoting the full inclusion and active participation in society of persons with disabilities, along with all citizens, through sustainable ICT solutions. Moreover, special attention is dedicated to the key role played by accessible digital technology in actively involving relevant stakeholders in a mutual exchange of opinions and information with the scientific community, thus creating new ways to communicate and interact between society and science to successfully address effective societal needs.

The development of the digital tool named “SmartInclusion@unipv” (SI@), which is one of the outcome of the project, is presented as a case study of sustainable ICT.

2.1 The App SmartInclusion@unipv (SI@) as a case study

“SmartInclusion@unipv” (SI@) is a smartphone app working on iOS which puts into operation an innovative, inclusive, sustainable and economic digital technology within the ancient courtyards of the University of Pavia. Thanks to a set of Bluetooth beacons (a class of Bluetooth low energy devices that broadcast their identifier to nearby portable electronic devices) positioned within the courtyards of the Palazzo Centrale of the University, the app enables students and researchers with blindness or low vision to move independently in areas where both GPS and Wi-Fi connectivity is very poor or absent due to structural/architectural limitations of the ancient building.

To date, the digital tool SI@ has been developing as a pilot project under a first phase of

implementation with basic functionalities under the research project «Building an Inclusive Digital Society for Vulnerable Persons: The Role of Social Media Tools in a Disability Human Rights Perspective» (hereafter, project BIDS), financed by the University of Pavia under the call Blue Sky Research Project 2017. The app in the pilot project version works in iOS, as this is the operative system commonly used among persons with blindness or low vision due to its accessibility features.

However, SI@'s functionalities will be further implemented during the project RISID in order to provide multiple services. In the first instance, new pathways will be developed adding numerous beacons to cover the main points of interests of the Palazzo Centrale of the University. In addition, new functionalities will be implemented on the basis of the preferences collected through surveys and brainstorming studies.

The ambition is to provide multiple services and tackle the need of different users – with and without disabilities – passing by the University's courtyards. In this respect, the digital tool could be developed to meet the needs of students (information on the location and timetable of classes, exams, libraries and other services), visiting researchers (information on the location of venues of conferences), tourists (artistic and historical information of the courtyards), and the whole community (information improving the safety, such as alerts on the work in progress within the courtyards, on barriers and eventual deviations on the journey). Finally, a version of SI@ compatible with Android will be released, in order to ensure the major spread of the mobile app to all the potential users.

3. Conclusion

The App SI@ will be used to implement the accessibility of the courtyards of the Palazzo Centrale of the University of Pavia and to effectively share scientific knowledge, thus communicating in a simple and captivating way – through a mobile app – the research activities carried out into the University. The mutual exchange of ideas and synergies between the society and the scientific community will also be ensured by implementing surveys and questionnaires via the app.

At the end of the project RISID, this tool will remain available for all researchers, students and the local community, as it will be used by the University as an asset to connecting the scientific community with society through a mobile app, giving free accessibility to physical and digital new spaces

References

- Ellis K., Kent M. (2016), *Disability and social media: Global perspectives*, Taylor & Francis.
- Foley A., Ferri B.A. (2012), *Technology for people, not disabilities: ensuring access and inclusion*, *Journal of Research in Special Educational Needs* (12) pp 192 ss..
- Lazar J., Stein M.A. (2017), *Disability, human rights, and information technology*, Philadelphia University Press.
- Ricci C. (2019), *Building an Inclusive Digital Society for Persons with Disabilities. New Challenges and Future Potentials*, Pavia University Press.

Wilsdon J., Willis, R. (2014), See-through science: Why public engagement needs to move upstream, Demos.

Wynne B. (2011), Public engagement as means of restoring trust in science? Hitting the notes, but missing the music, Community Genetics, 2006, pp 211 ss

Authors

Silvia Favalli - silvia.favalli@unipv.it

LPost-doctoral research fellow (assegnista di ricerca) in International Law at the Department of Political and Social Sciences of the University of Pavia; P.I. of the project RISID

Flavio Ceravolo - flavioantonio.ceravolo@unipv.it

Assistant Professor in Sociology at the Department of Political and Social Sciences of the University of Pavia

Matilde Ceron - matilde.ceron@unipv.it

Post-doctoral research fellow (assegnista di ricerca) in Economics at the Department of Political and Social Sciences of the University of Pavia

Anna Facchinetti - anna.facchinetti@unipv.it

Research fellow (borsista di ricerca) in International law at the Department of Political and Social Sciences of the University of Pavia

Giulia Melis - giulia.melis@unipv.it

Research fellow (borsista di ricerca) in Sociology at the Department of Political and Social Sciences of the University of Pavia

Vittorio Pampanin - vittorio.pampanin@unipv.it

Fixed-term research assistant (RTD-A) (ricercatore) in Administrative Law at the Department of Law of the University of Pavia

Marco Porta - marco.porta@unipv.it

Associate Professor at the Department of Electrical, Computer and Biomedical Engineering of the University of Pavia

Beatrice Rabai - beatrice.rabai@unipv.it

Research fellow (borsista di ricerca) in Administrative Law at the Department of Law of the University of Pavia

Carola Ricci - carola.ricci@unipv.it

Associate Professor in International Law at the Department of Political and Social Sciences of the University of Pavia

Francesco Scervinii - francesco.scervini@unipv.it

Associate Professor of Economic Policy at the Department of Political and Social Sciences of the University of Pavia

TrashWare e Digital Divide

Alexander Palummo

Informatici Senza Frontiere

Abstract. Il contributo analizza il rapporto tra le varie forme di esclusione dal modo dell'informatica ricomprese nel divario digitale e le potenzialità (anche di contrasto a tale divario) del recupero di hardware obsoleto meglio noto come trashware. Le implicazioni del recupero hardware sono principalmente etiche, ambientali ed economiche ma essendo un'attività ancora poco diffusa viene difficilmente abbinata alle (più diffuse) attività di contrasto al divario digitale. Si propone pertanto di ovviare a tale problema a partire dalla maggiore diffusione di conoscenze e prassi in materia, a partire proprio dai contesti istituzionalmente preposti all'educazione e alla formazione

Keywords. TrashWare, Digital Divide, HardWare, obsolescenza programmata

Introduzione

Con divario digitale (o digital divide) si intende il gap tra quanti hanno accesso alle tecnologie informatiche (hardware, software e di connettività) e quanti per motivi economici, culturali, etnici, sanitari ecc. non ce l'hanno. Si discute di divario digitale da oltre trent'anni: ma mentre ad oggi sul piano dell'uso del software e dell'accesso a Internet si è creato un dibattito intenso (seppur, talvolta, un poco di nicchia), per quanto riguarda l'hardware il discorso non si è evoluto di pari passo. Questo ritardo è problematico perché le sue conseguenze sociali, economiche, sanitarie e ambientali concorrono ad alimentare un altro e più antico gap – quello tra Nord e Sud del mondo – che nell'attuale contesto di saturazione del commercio globale e di esaurimento delle risorse economiche e ambientali rischia di tornarci minacciosamente indietro.

1. Che cos'è il TrashWare

Con trashware ci riferiamo all'attività di recupero di materiale informatico normalmente destinato alla spazzatura (in inglese trash). Materiale hardware spesso datato, ancora funzionante (o comunque riutilizzabile o aggiustabile), talora proveniente da computer diversi, viene ri-assemblato per dare vita a macchine ricondizionate e destinabili anche a scopi di utilità sociale, come l'alfabetizzazione informatica e altre attività a contrasto del divario digitale.

Il software libero, più "leggero" e più adattabile alle esigenze di macchine eterogenee, concorre al recupero di tale materiale. Infatti è solo il connubio tra il FreeSoftware e il TrashWare che consente alfabetizzazione informatica e riduzione del digital divide a basso impatto ambientale e nel rispetto della legalità e della gratuità delle licenze d'uso. Questo grazie a:

- facile reperibilità del software stesso (libero download dal web);
- trasparenza e immediatezza della didattica (libertà di studiare e modificare il software)

- (R. Stallman, 2003);
- affidabilità e qualità tecnica (a partire dal sistema operativo e le sue interfacce) (E. S. Raymond, 1998);
 - libera concorrenza e continuo stimolo della conoscenza (“Posso farlo anch'io!”) (A. R. Meo e M. Berra, 2001);
 - vantaggio economico (prevalente gratuità) ed ecologico (risparmio energetico per l'uso di sistemi meno prestanti o ottimizzati per essere prestanti con meno risorse).

Il trashware può avere punti di contatto con il retrocomputing (il recupero di macchine datate e non modificate con finalità prevalentemente collezionistiche), il modding (la personalizzazione dell'hardware, in genere il contenitore, per scopi solitamente estetici) o ancora il refurbishment (il mantenimento o la riparazione estetica o meccanica di componenti per lo più destinate alla re-immissione nel mercato a un prezzo più contenuto rispetto ai loro equivalenti venduti come nuovi). Ma, a differenza di queste ultime pratiche, il ricorso al trashware ha sempre implicazioni morali, ecologiche e di solidarietà internazionale che si integrano con alcuni elementi da tempo attenzionati dall'ethics of IT (M. Berra e A. R. Meo, 2006):

- divario digitale;
- open source e open data;
- obsolescenza programmata;
- domanda di tecnologia (dispositivi e banda);
- salute/sanità pubblica.

2. La “spazzatura” informatica

La normativa cui di solito si fa capo in materia di rifiuto elettronico (direttiva 2012/19/UE) non basta per comprendere il problema dei rifiuti informatici (che sono solo una parte dei RAEE).

Il mondo dell'electronic and electrical engineering (EEE), richiede cicli di raccolta e smaltimento dedicati già in fase di produzione delle macchine. Ma quando queste si trasformano e-Waste propriamente detto, una parte di esse verrà recuperata e riutilizzata come materia prima (oro, argento, rame e altri metalli riassorbibili nel circuito di economia circolare delle cosiddette “miniere urbane”). Un'altra parte, non recuperabile e spesso non inerte, va invece assolutamente raccolta, tracciata e trattata: si tratta di materiale spesso cancerogeno che è oneroso da stoccare o smaltire nel rispetto delle normative ambientali tipiche di paesi a economia avanzata come quelli europei.

Fig. 1
Possibili conseguenze sulla salute a seguito della dispersione di materiali tossici presenti all'interno di un computer
(V. Forti et al., 2020)

Materiali Tossici	Malformazioni congenite	Danni al cervello	Danni a cuore, milza, polmoni e fegato	Danni ai reni	Danni al sistema nervoso e riproduttivo	Danni al sistema scheletrico	Esempi dove possiamo trovare questo elemento
Bario (Ba)		x	x				Monitor CRT
Cadmio (Cd)	x		x	x	x	x	Chip
Piombo (Pb)	x	x		x	x		Saldature
Litio (Li)	x	x	x	x	x		Batterie
Mercurio (Hg)	x	x	x	x	x		Circuiti
Nichel (Ni)	x		x	x	x		Cavi
Palladio (Pd)	x	x	x	x			Vetri e pellicole
Rodio (Rh)			x				Contatti elettrici
Argento (Ag)	x	x	x	x	x		Contatti elettrici
Cromo (Cr)	x		x				Nastri magnetici
Berillio (Be)			x	x	x		Schede madri
Arsenico (As)					x	x	Plastiche

Per questo sono frequenti le soluzioni “abusive” in cui pezzi di hardware gettati nell’indifferenziata o abbandonati in discariche improvvisate contaminano irrimediabilmente l’ambiente circostante e gli esseri umani o animali che vivono intorno (ad esempio, il tubo catodico dei vecchi monitor se rotto può rilasciare sostanze tossiche e/o cancerogene).



Fig. 2
Mappa dei flussi di rifiuto informatico nel mondo durante l'anno 2019 (J. Okeme e V. H. Arrandale, 2019)

Purtroppo, però, anche il ciclo di smaltimento regolare ha dei limiti di tracciabilità: fuori dal mondo occidentale, il materiale considerato non recuperabile viene accumulato nelle regioni più povere e remote del mondo in improvvisate discariche a cielo aperto dalle dimensioni impressionanti, in cui vivono e lavorano uomini, donne e bambini assolutamente ignari dei seri problemi di salute in cui in cui incorrono quotidianamente. Tali aree di raccolta di e-Waste, complice una debole normativa locale in materia di gestione dei rifiuti tossici, non sono nemmeno vagamente attrezzate per il trattamento del rifiuto informatico, con danno ambientale e per la salute umana stimato come ingente e spesso irreversibile.

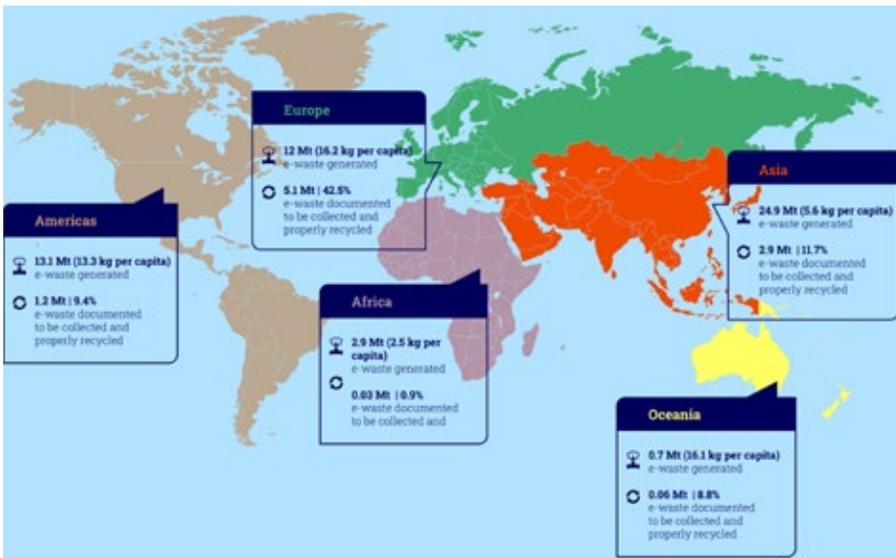


Fig. 3
Mappa che mette in rapporto la produzione e il riciclo di rifiuto elettronico per continente durante l'anno 2019 (V. Forti et al., 2020)

3. Un nesso con il Digital Divide

Il divario digitale, inteso come gap tra chi ha accesso all’informatica e chi no, può avere declinazioni economiche, di genere, etiche, culturali, anagrafiche ecc. e manifestarsi in vari

modi (dal mancato accesso all'Internet all'incapacità di mantenere un dispositivo). Confrontando la mappa in Figura 2 con la mappa in Figura 4, emerge che, tendenzialmente, i paesi destinatari di grandi quantità di e-Waste e i paesi a maggior divario digitale coincidono.



Fig. 4
Mappa estratta da un articolo di government.no che mostra la percentuale di popolazione che ha accesso a Internet nel 2019

Sottrarre alla spazzatura l'hardware dismesso ci aiuta a non essere più complici di una grave ingiustizia nei confronti dell'uomo e dell'ambiente e a improntare il nostro stile di vita a una maggiore sobrietà. La formazione scolastica o comunque istituzionale occidentale avrebbe al contempo la possibilità e la responsabilità di includere questi aspetti all'interno dei propri programmi di studio. Purtroppo, però, anche ai livelli più avanzati del sapere queste nozioni – che pure darebbero un valido contributo in termini di risparmio di risorse naturali ed economiche e di valorizzazione dell'equità sociale – tendono a restare appannaggio di pochi addetti ai lavori.

4. Conclusioni

Nell'epoca dell'economia circolare una riflessione sulla trasformazione dei rifiuti in materie prime non può trascurare il mondo del rifiuto elettrico ed elettronico, sia per i suoi volumi economici e materiali, sia per le sue potenziali conseguenze sociali ed ambientali. La dislocazione geografica delle materie prime (litio, terre rare, ecc.) e degli impianti di produzione dei nostri oggetti elettronici di consumo ci hanno abituati all'origine remota del nostro smartphone o del nostro tablet. Ma il fatto che questa condizione venga ormai data per scontata non ci esime dal porci certe domande, soprattutto una volta appresi dati inquietanti sulle conseguenze etiche, ambientali ed economiche dell'invio dei nostri dispositivi in luoghi altrettanto remoti sotto forma di e-Waste.

Alla responsabilità sociale ed economica verso i paesi più poveri da cui preleviamo le materie prime e dei quali sottopaghiamo la manodopera che produce dispositivi elettronici (cosa che alimenta, tra le tante iniquità locali, anche il divario digitale) si aggiunge ora quella sanitaria e ambientale.

Per l'una e per l'altra in molti casi non si può più parlare di reversibilità delle conseguenze

e, se comprendiamo quanto siamo interconnessi con il resto del Pianeta, non possiamo non improntare il nostro stile di vita a modelli di consumo più sobri. E il trashware è sicuramente tra questi.

Riferimenti bibliografici

- A. R. Meo, M. Berra. (2001), *Informatica solidale: storia e prospettive del software libero*, Bollati e Boringhieri, Torino.
- C. Baldé, V. Forti, V. Gray, R. Kuehr, P. Stegmann. (2017), *The global e-waste monitor 2017*, United Nations University (UNU), International Telecommunication Union (ITU) & International Solid Waste Association (ISWA), Bonn/Geneva/Vienna.
- C. P. Baldé, R. Kuehr, K. Blumenthal, S. Fondeur Gill, M. Kern, P. Micheli, E. Magpantay, J. Huisman. (2015), *E-Waste Statistics Guidelines on Classification, Reporting and Indicators*.
- D. Della Porta, L. Mosca. (2003), *Globalizzazione e movimenti sociali*, Roma.
- E. S. Raymond. (1998), *La cattedrale e il bazar*, Apogeo.
- J. Okeme, V. H. Arrandale. (2019), *Electronic Waste Recycling: Occupational Exposures and Work-Related Health Effects*. *Curr Envir Health Rpt* 6, 256–268.
- L. De Biase, P. Soldavini. (2013), *Atlante Geopolitico 2013 Treccani*.
- M. Berra, A. R. Meo. (2006), *Informatica solidale 2. Libertà di software, hardware e conoscenza*, Bollati e Boringhieri, Torino.
- M. Masini, A. Lovari, S. Benenati (a cura di). (2011), *Dal digital divide ai media sociali. Le nuove sfide della comunicazione pubblica*, Torino.
- P. Aigrain. (2012), *Sharing. Culture and Economy in the Internet Age*, Amsterdam.
- R Stallman. (2003-04), *Software Libero, Pensiero Libero* (vol.1 e 2), Stampa Alternativa.
- S. Bentivegna. (2009), *Disuguaglianze digitali. Le nuove forme di esclusione nella società d'informazione*, Roma-Bari.
- S. Rodotà. (2004), *Tecnopolitica. La democrazia e le nuove tecnologie della comunicazione*, Roma-Bari.
- V. Forti, C. P. Baldé, R. Kuehr, G. Bel. (2020), *The Global E-waste Monitor 2020*, UNU/UNITAR and ITU.
- www.regjeringen.no/en/dokumenter/meldst11_summary/id2699502/?ch=2 (2019)
- wearesocial.com, Norvegia.

Autore

Alexander Palummo alexander.palummo@informaticisenzafrontiere.org

Ph.D. in Pianificazione Territoriale e appassionato di informatica, soprattutto hardware, individua nel trashware il punto di contatto tra tale passione e le altrettanto importanti istanze ecologiste di cui, anche in altri contesti, si fa portatore. Con Informatici Senza Frontiere si è occupato di corsi sia di alfabetizzazione informatica che di trashware. Socio da diversi anni di ISF, è dal gennaio 2018 coordinatore di ISF Toscana e si occupa anche di progetti di cooperazione internazionale.

A brand-new flexible and scalable ENEA solution to handle data in smart city environment

Giuseppe Santomauro, Marta Chinnici, Giovanni Ponti

ENEA

Abstract. In this work, the authors propose an approach and describe a methodology and show a modular and scalable multi-layered ICT platform called ENEA Smart City Platform (ENEA-SCP). The solution provided by ENEA SCP to exploit potentials in Smart City environments is based on four fundamental concepts: Open Data, Interoperability, Scalability and Replicability. The ENEA-SCP is implemented following the Software as a Service (SaaS) paradigm, exploiting cloud computing facilities to ensure flexibility and scalability. Interoperability and communication are addressed employing web services, and data format exchange is based on the JSON data format. By taking into account these guidelines as references, this talk provides a description of the SCP developed by ENEA and its potential use for smart and IoT city applications

Keywords. Smart City, Big Data, ICT Platform, Interoperability, IoT

Introduction

In a smart city environment, the explosive growth in the volume, speed, and variety of data being produced every day requires a continuous increase in the processing speeds of servers and entire network infrastructures, platforms as well as new resource management models. This poses significant challenges (and provides attractive development opportunities) for data-intensive and high-performance computing, i.e., how to turn enormous datasets into valuable information and meaningful knowledge efficiently. The variety of sources complicates the task of context data management such as data derives from, resulting in different data formats, with varying storage, transformation, delivery, and archiving requirements. At the same time, rapid responses are needed for real-time applications. With the emergence of cloud infrastructures and platforms, achieving highly scalable data management in such contexts is a critical problem, as the overall urban application performance is highly dependent on the properties of the data management service. This means, continuously developing and adopting ICT technologies to create and use platforms for government, business and citizens can communicate and work together and provide the necessary connections between the networks that are the base for the services of the smart city (Chinnici et al, 2018].

The main features of a generic Smart City Platform (SCP) are in the following (Brutti et al., 2018).

- Make data, information, people and organizations smarter;
- Redesign the relationships between government, private sector, non-profits, communities and citizens;

- Ensure synergies and interoperability within and across city policy domains and systems (e.g. transportation, energy, education, health & care, utilities, etc.);
- Drive innovation, for example, through so-called open data, living labs and tech-hub.

1. ENEA Smart City Platform

The ENEA-SCP is implemented following the Software as a Service (SaaS) paradigm, exploiting cloud computing facilities to ensure flexibility and scalability. Interoperability and communication are addressed employing web services, and data format exchange is based on the JSON data format. By taking into account these guidelines as references, this work provides a description of the SCP developed by ENEA and its potential use for smart and IoT city applications.

The solution provided by ENEA SCP to exploit potentials in Smart City environments is based on four fundamental concepts:

- Open Data
- Interoperability
- Scalability
- Replicability

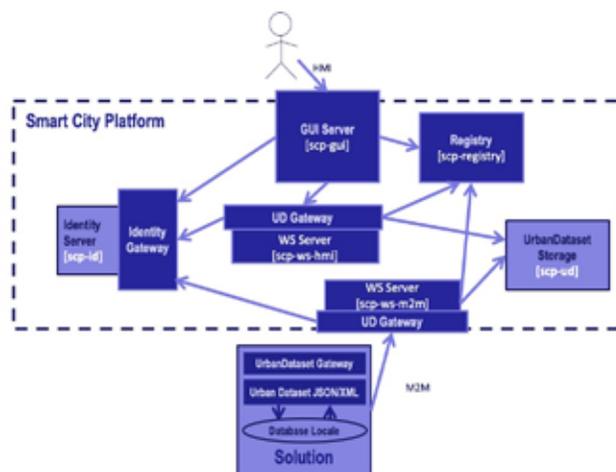
In this scenario, the ENEA SCP is going to tackle the issues concerning these two aspects providing a reference framework of modular (Brutti et al., 2018) specifications for stakeholders willing to implement ICT platforms to exploit the Smart City vision potentials and therefore offer new services for the citizen.

The ENEA Smart City Platform exploits computational resources of the ENEAGRID infrastructure (Iannone et al., 2019), as it is deployed in the cloud hosted in the Portici Research Center site. The creation of a customized environment ENEA cloud-based platform is possible thanks to the virtualization technologies of VMWARE platform, which allows hosting the management, the transportation and the processing of project data services, ensuring their availability and protection over time. More in detail, the SCP is composed by six Virtual Machines (VMs), and each of them hosts a component with a specific role (Fig. 1).

2. Conclusions

In this work, the authors presented the interoperability topic of Smart City platforms and attempted to provide their approach presented by the ICT ENEA Smart City Platform (SCP). This platform is able to interact with different stakeholders at the urban district level by a flexible and multipurpose data format. The ENEA-SCP milestones are the op-

Fig. 1
Scheme of
ENEASCP
architecture
based on VMs



portunity to scale the computational resources according to requests; interoperate with all the interesting parts; and replicate all the components in the different city context.

References

Brutti A., De Sabbata P., et al., (2018), Smart City Platform Specification: A Modular Approach to Achieve Interoperability in Smart Cities, Chapter book in The Internet of Things for Smart Urban Ecosystems pp 25-50, Springer.

Chinnici M., De Vito S., (2018), IoT Meets Opportunities and Challenges: Edge Computing in Deep Urban Environment, in Dependable IoT for Human and Industry. Modeling, Architecting, Implementation. River Publishers Series in Information Science and Technology.

Iannone F. et al., (2019), "CRESCO ENEA HPC clusters: a working example of a multi-fabric GPFS Spectrum Scale layout," 2019 International Conference on High Performance Computing & Simulation (HPCS), Dublin, Ireland, pp. 1051-1052, doi: 10.1109/HPCS48598.2019.9188135.

Authors



Giuseppe Santomauro giuseppe.santomauro@enea.it

Giuseppe Santomauro is a researcher at HPC-Lab of the ENEA Research Center in Portici (NA). He won an scholarship at the Consortium GARR. He holds a PhD in Mathematics and Computer Sciences at the University of Naples. He holds a degree in Mathematics at the University of Salerno. Its main activities concern: HPC systems management; research on ICT environments for Smart Cities; crawling and analysis of data from the Web; efficient numerical methods development for high performance computing.

Marta Chinnici marta.chinnici@enea.it

Marta Chinnici received the degree in mathematics and the Ph.D. degree in mathematics and computer science from the University of Naples. She is a Senior Researcher with the ICT Division at ENEA Casaccia Research Center. She works on ICT for Energy Efficiency issue. She is a European Commission Expert as review/evaluator in the field of ICT for European programs. She is a member of technical programme committee of various international conferences and workshop and an editor/referee of relevant journals.



Giovanni Ponti giovanni.ponti@enea.it

Giovanni Ponti has been a researcher at ENEA since December, 2010, where he is head of Computer Science Management System Laboratory since August, 2020. He is member of Technical-Scientific Committee at GARR since July 2018. He graduated magna cum laude in Computer Engineering at the University of Calabria, in 2005, and obtained his Ph.D. in Computer Engineering, in 2010. In 2010, he worked as a post-doc at the DEIS Department, focusing on information and knowledge extraction from complex data, with particular interest in data modeling and data mining algorithms. His research interests concern Cloud Computing, Big Data e Data Analytics. On these topics, he has coauthored journal articles, conference papers, book chapters, and technical reports. He also developed practical software tools. He has served as a reviewer for leading journals and conferences in the field of databases and data mining.

Ludoteca del Registro .it: i laboratori di cybersecurity

Giorgia Bassi, Beatrice Lami

CNR-IIT, Istituto di Informatica e Telematica

Abstract. Il contributo descrive i laboratori di cybersecurity della Ludoteca del Registro .it, progetto a cura del Registro .it, anagrafe dei domini internet italiani il cui servizio è gestito dall'Istituto di Informatica e Telematica del Cnr di Pisa

Keywords. Digital literacy, cybersecurity, educazione digitale

Introduzione

La Ludoteca del Registro .it è un progetto nato nel 2011 con l'obiettivo di diffondere la cultura digitale a partire dalle scuole primarie. Ad oggi, sono oltre 14.000 gli alunni incontrati in tutto il territorio nazionale, per un totale di circa 1400 ore di formazione.

Il progetto nasce da un dato evidente: ragazzi e bambini trascorrono buona parte della loro vita online, nella maggioranza dei casi senza gli strumenti adeguati per un utilizzo sicuro e responsabile delle risorse digitali.

Secondo la ricerca "Ipsos per Save the Children" del 2019, il 97% dei ragazzi tra gli 11 e 17 anni possiede uno smartphone connesso a Internet e il 74% dichiara di avere una percezione dei potenziali rischi. A questo si aggiunge il ruolo della famiglia, non sempre preparata a gestire la dimensione digitale della vita dei figli. Secondo la ricerca condotta da CISF (2017) su un campione rappresentativo di 4000 famiglie, il 56% dichiara di non avere protocolli condivisi per l'uso dei media digitali, il 27% non ne parla mai con i figli e il 37% non si preoccupa di quello che questi ultimi possono pubblicare online. Inoltre, nella maggior parte dei casi "a fronte di una verbalizzazione preoccupata dei rischi del digitale, si assiste a una diffusione generalizzata e non adeguatamente resa oggetto di riflessione critica" (Rivoltella 2020).

I problemi derivanti da un uso non sicuro di Internet e dei sistemi informatici possono essere molteplici, dalla diffusione di malware, alla violazione di dati al furto di identità. Fenomeni sempre più diffusi come il cyberbullismo, il grooming e il sexting implicano invece considerazioni e interventi che appartengono all'ambito delle scienze sociali, ma è evidente che gli strumenti della cybersecurity, anche in questi casi, possono rappresentare un primo passo per stimolare un atteggiamento preventivo, soprattutto per quanto riguarda la protezione dei dati personali.

1. Insegnare la cybersecurity

A partire dall'anno scolastico 2018/19, la Ludoteca del Registro .it ha avviato il progetto di

laboratori di cybersecurity per le classi primarie e secondarie di primo grado, coinvolgendo in questo primo ciclo circa 1000 bambini.

I laboratori prevedono una parte propedeutica dedicata a nozioni elementari di informatica (ad esempio: indirizzo IP, protocolli informatici, nomi a dominio) spiegate sempre attraverso attività di gruppo. Si passa quindi a introdurre la cybersecurity come insieme degli strumenti per la protezione del cyberspazio, intendendo con questo termine l'insieme dei sistemi informatici (computer e dispositivi digitali di vario tipo) e dei dati in formato digitale.

Ai bambini si spiega che la sicurezza, in generale, va intesa come un processo, non come una soluzione tecnica da usare per contrastare la minaccia quando si presenta: anche in ambito informatico, è fondamentale sviluppare un atteggiamento preventivo. A questo proposito, le immagini di una cintura di sicurezza e di uno spazzolino da denti risultano molto efficaci per introdurre il concetto di "igiene informatica", l'insieme cioè dei comportamenti da seguire quotidianamente per prevenire e minimizzare i rischi.

La difesa del cyberspazio riguarda non solo l'aspetto hardware dei dispositivi ma anche i dati in essi salvati, trasmessi e condivisi. Per introdurre il concetto di dati personali e privacy in Rete, viene mostrata l'immagine di un cartello di proprietà privata, segnale che i bambini riconoscono e che identificano come divieto di accesso all'interno di uno spazio fisico.

Argomento quello della privacy online cruciale visto che, per la cosiddetta generazione Z (i nati dalla seconda metà degli anni '90 al 2010), rappresenta un valore tutto da definire, la "self(ie) generation" vive in "un contesto di sovra esposizione di informazioni personali, si costruisce attraverso le proprietà che caratterizzano il rapporto tra relazioni e contenuti nelle proprie reti, aprendo o chiudendo le cerchie, esplorando i profili degli altri a misurarne la reputazione, dando valore a sharing e tag" (Boccia Artieri 2015).

Un concetto quindi molto labile che merita di essere chiarito già nelle classi primarie, definendo quali siano le informazioni che senza dubbio non si devono mai condividere, ovvero password, nominativi, indirizzi postali, numeri di telefono, dati bancari.

Ai bambini si spiegano quindi i principali requisiti di un sistema informatico sicuro: confidenzialità, integrità e disponibilità. La confidenzialità è la riservatezza di un messaggio (si fa l'esempio di una mail o di una chat), l'integrità è la completezza e correttezza di quel messaggio (si pone la domanda: che cosa succederebbe se al destinatario arrivassero messaggi interrotti a metà?), la disponibilità è la caratteristica di rendere quel messaggio o quell'applicazione leggibile/utilizzabile per tutto il tempo necessario.

2. Conoscere e contrastare le minacce

Il tema delle minacce e della difesa del cyberspazio è affrontato portando all'attenzione della classe esperienze e opinioni, attraverso queste domande poste alla classe: "avete mai preso un virus informatico o è capitato a qualcuno che conoscete?", "come si deve impostare una password perché sia sicura?", "avete mai sentito parlare di "hacker", "quali sono i dati che non si devono condividere su Internet?".

Attraverso il dibattito, si stimola una prima riflessione su una serie di comportamenti

a rischio: apertura e condivisione di link all'interno di messaggi che provengono da sconosciuti, scelta inadeguata di password, download applicazioni da siti non sicuri, condivisione dati personali, gestione inadeguata dei permessi delle app e delle impostazioni di accesso e privacy di account social o di applicazioni di messaggistica istantanea.

L'immagine utilizzata per approfondire il tema delle minacce e degli attacchi cyber è il cavallo di Troia che, oltre a indicare una specifica categoria di malware, permette ai bambini di riflettere soprattutto sui meccanismi di diffusione.

Quello che si evidenzia è che l'anello debole è rappresentato, proprio come nella leggenda dell'assedio di Troia, non tanto da eventuali vulnerabilità tecniche ma soprattutto dal comportamento umano che cede, per esempio, di fronte a invitanti premi o richieste da parte di persone che si fingono amici.

In questo caso, l'atteggiamento da adottare è quello di una "sana diffidenza" che deve passare anche dal confronto e dalla condivisione delle esperienze con gli adulti, spesso meno esperti da un punto di vista tecnico ma sicuramente in grado di riconoscere e gestire alcune situazioni di potenziale rischio.

Dopo una breve introduzione ai più comuni malware (virus, worm, trojan, spyware) e soprattutto ai meccanismi che permettono a questi programmi malevoli di entrare nei dispositivi (link sospetti, download da siti non ufficiali), si affronta il tema delle contromisure.

Tra queste i sistemi di autenticazione, in particolare ci si sofferma sul tema delle password, giocando con una serie di combinazioni per dimostrare l'importanza di sceglierle sempre robuste, ovvero non facili da memorizzare e diverse a seconda degli account. Introduciamo quindi altri sistemi di difesa come l'antivirus e i firewall con un breve accenno anche alla crittografia che approfondiamo nel gioco del Cifrario di Cesare, descritto nel paragrafo successivo.

3. Giocare con la sicurezza informatica

La caratteristica comune dei laboratori è quella di alternare le nozioni teoriche con attività pratiche di gioco che vedono la partecipazione attiva dei bambini. Per il dettaglio completo dei materiali utilizzati si rimanda al sito web della Ludoteca, alla pagina: <https://www.ludotecaregistro.it/per-le-scuole/cybersecurity/>.

Di seguito una breve descrizione di alcuni materiali utilizzati:

- Fumetto "Nabbovaldo contro i pc zombi": ambientato nella città di Internetopoli, il protagonista è Nabbovaldo, un ragazzo appassionato di Internet ma molto ingenuo che dovrà affrontare la minaccia di terribili malware.
- Gioco del "Cyber Security Quiz": tavole a fumetti ambientate nella città di Internet. Il protagonista è sempre Nabbovaldo. Ogni tavola presenta una situazione a rischio di partenza e tre possibili comportamenti che potrebbero risolverla, ma solo uno rappresenta la scelta corretta.
- Cyber bowling sulla sicurezza in Rete: i birilli da abbattere sono i comportamenti da evitare, come ad esempio non aggiornare i sistemi operativi, scegliere sempre la stessa password, scaricare app da siti non ufficiali.

Fig. 1
Il gioco a fumetti
"Cyber Security Quiz"



- Cifrario di Giulio Cesare: l'antico strumento per inviare messaggi segreti diventa lo spunto per spiegare ai bambini la confidenzialità dei dati e una tecnica di crittografia.
- Password Memory: lo sforzo mnemonico legato alla ricerca di coppie uguali di password è un'occasione per riflettere sull'importanza di sceglierle in modo adeguato.

Naturalmente, a seguito dell'emergenza da Covid-19, durante l'anno scolastico 2020/21, si è resa evidente l'importanza di adottare strumenti per la didattica a distanza e dunque alcune attività sopra ricordate sono state in parte adattate al mutato contesto di apprendimento.

4. Sviluppi futuri

Nel mese di maggio 2021 è stato pubblicato sui principali store il videogioco "Nabbovaldo e il ricatto dal cyberspazio" per dispositivi mobile e nella versione desktop. La scelta di sviluppare, all'interno di questo progetto, un videogioco educativo interamente dedicato a questo tema rappresenta un'ulteriore spinta nella diffusione di una cultura della sicurezza informatica attraverso una modalità didattica innovativa e ricca di potenzialità.

La modalità di fruizione è single-player ma è previsto un percorso specifico di utilizzo nelle classi, attraverso il supporto dei formatori della Ludoteca e degli altri strumenti e risorse del progetto. Il gioco prevede una struttura ibrida tra il percorso fisso e open world: ci si può muovere liberamente nella mappa della città, parlare con i personaggi e risolvere i mini-giochi nell'ordine che si preferisce, anche se la trama del gioco si sviluppa in quattro capitoli principali, più un epilogo.

Dal mese di febbraio 2021 inoltre è stata avviata la valutazione dell'efficacia dei laboratori in collaborazione con il Dipartimento di Formazione, Lingue, Intercultura, Letterature e Psicologia dell'Università di Firenze.

La valutazione si è svolta fino al mese di aprile 2021 mediante la somministrazione di questionari ex ante ed ex post.

Riferimenti bibliografici

Boccia Artieri G., (2015), Gli effetti sociali del Web. Forme della comunicazione e metodologie della ricerca online, Franco Angeli, Milano.

Rivoltella P. C., (2020), Nuovi alfabeti, Scholè, Brescia.

Autrici



Giorgia Bassi 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. E' referente del progetto di peer education Let's Bit! destinato agli istituti superiori.

Beatrice Lami 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!.



Cybersecurity: sfide e prospettive per una società più sostenibile

Angelica Maiorano

Università degli Studi di Bari "Aldo Moro"

Abstract. La rivoluzione tecnologica ha trasformato la vita delle persone. Contestualmente, si è affermato un concetto di sviluppo sostenibile quadridimensionale. Di conseguenza, l'innovazione tecnologica e lo sviluppo sostenibile costituiscono i principali obiettivi delle più recenti iniziative internazionali. Nazioni Unite e Unione europea hanno avviato programmi d'azione finalizzati ad avviare una crescita sostenibile fondata sullo sfruttamento del potenziale tecnologico a vantaggio della società. In questo contesto la cybersecurity è il presupposto per il perseguimento di simili obiettivi, data la necessità di garantire la sicurezza delle infrastrutture digitali per creare fiducia nello spazio online e incentivare l'utilizzo diffuso ed efficace delle tecnologie. A tal fine sono sempre più numerosi gli sforzi per costruire un dialogo multilaterale sulla pace cibernetica e rafforzare le reti informatiche, soprattutto quelle relative ai servizi cruciali per lo sviluppo sostenibile globale

Keywords. Cybersecurity, sostenibilità integrata, rivoluzione tecnologica, pace cibernetica, cooperazione internazionale.

Introduzione

La rivoluzione tecnologica ha trasformato la vita delle persone, rendendo di primaria importanza l'adozione di un approccio antropocentrico alla tecnologia, che ponga quest'ultima al servizio dei cittadini. La società contemporanea, infatti, si basa sui dati. Essi definiscono il modo in cui si produce, si consuma e, più in generale, si vive. Quotidianamente loro ingenti quantità vengono immesse nella rete digitale attraverso le moderne tecnologie di informazione e comunicazione (d'ora in poi ICT). La rilevanza è tale da porre gli stessi a fondamento dei più innovativi processi decisionali così da adeguare le decisioni alle esigenze e alle preferenze degli utenti, ottimizzando risorse e risultati.

Siffatto stato di cose ha due immediati corollari pratici: per un verso, sono consentiti rapidi scambi di informazioni funzionali al progresso della società nel suo complesso; per altro verso, sono moltiplicate le occasioni per perpetrare violazioni e abusi che minano la fiducia dei fruitori dei servizi digitali e, pertanto, compromettono la sostenibilità dello sviluppo. Da qui l'urgenza di un intervento coordinato dei principali attori globali affinché si realizzi un sistema di regole e controlli in grado di garantire, al contempo, un accesso equo ai dati e alle tecnologie e la cybersecurity, ossia la sicurezza delle infrastrutture digitali.

Dal canto suo, lo sviluppo sostenibile costituisce uno degli obiettivi delle principali azioni internazionali più recenti, attesa l'insostenibilità dell'attuale modello di produzione e crescita. Azioni internazionali che, sempre più frequentemente, interpretano il progresso tecnologico e lo sviluppo sostenibile come due facce della stessa medaglia e, quindi, ritengono opportuno

sfruttare le potenzialità delle tecnologie in termini di quantità e qualità delle informazioni trasmesse nonché di grado di precisione nel perseguimento degli obiettivi per creare società più sostenibili.

Al riguardo giova ricordare l'Agenda 2030 per lo Sviluppo sostenibile, un programma d'azione con cui i 193 Paesi membri delle Nazioni Unite hanno individuato gli obiettivi necessari per una crescita sostenibile, intendendo per tale uno sviluppo che tenga conto della dimensione economica, dell'inclusione sociale, della vitalità culturale e della tutela dell'ambiente, e il Decennio digitale appena inaugurato, che mira a fare dell'Ue una società digitale dinamica, affidabile e verde capace di tradurre le ambizioni digitali in termini di concreto vantaggio per cittadini, imprese e istituzioni.

1. Cosa significa sostenibilità?

Alla luce di quanto appena esposto, è possibile affermare che l'innovazione tecnologica è strumentale al perseguimento di una crescita sostenibile. Tuttavia, affinché sia possibile comprendere pienamente la portata di siffatta correlazione e il fondamentale ruolo che la cybersecurity riveste in tale contesto, occorre preliminarmente analizzare il concetto di sviluppo sostenibile.

Secondo l'orientamento più recente, esso ha natura quadridimensionale, concorrendo alla sua definizione quattro sfere d'azione: responsabilità ambientale, equità sociale, dinamismo culturale e sicurezza economica (Benitez D. A. et al., 2019). Da siffatta impostazione discendono differenti modelli comportamentali che descrivono l'atteggiarsi della cittadinanza sostenibile.

Tra questi modelli, particolare rilevanza è assunta dallo schema affermatosi in dottrina che, individuato nel benessere l'elemento essenziale cui aspira ciascun consociato, indica le risposte alle domande che dovrebbero ispirare la vita di ogni persona. L'archetipo di cittadino che ne deriva è quello di un cittadino conscio di vivere in un contesto globale e interconnesso caratterizzato da ecosistemi e da processi sociali e tecnologici, nei confronti dei quali deve porsi come agente di valorizzazione, cambiamento e progresso; la sua aspirazione ultima deve essere la creazione di una società aperta, equa, tollerante e prospera. Ciononostante, l'analisi del moderno modello di produzione evidenzia la sua lontananza da tali principi con una scarsa valorizzazione della diversità e delle nuove conoscenze.

Lo stato delle cose, dunque, impone di adottare un sistema di sostenibilità c.d. integrata, che riconosca l'interconnessione tra tutti gli organismi e gli individui del pianeta e, pertanto, promuova processi di cooperazione globale (Fabiotti G. et al., 2014). Questo è il motivo per cui l'Obiettivo 17 dell'Agenda 2030 è incentrato sul rafforzamento dei mezzi di attuazione del programma e sulla rinnovazione del partenariato multilaterale quale imprescindibile strumento di collaborazione per lo sviluppo sostenibile. Tra le azioni necessarie per il suo perseguimento figura la creazione di un meccanismo globale di accesso alla tecnologia, che consenta, tra le altre cose, un uso maggiore e sicuro delle tecnologie avanzate.

La sostenibilità integrata è la ragione, altresì, per cui il Decennio digitale europeo si pone come fine la creazione di un ambiente online sicuro e sostenibile grazie a progetti che combinino investimenti pubblici e privati, europei e di Stati terzi, promuovendo così una

cooperazione multilivello.

2. Azioni di cybersecurity per un mondo digitale più sicuro

Quanto descritto finora ha messo in luce la centralità della fiducia nei processi tecnologici e sostenibili: la digitalizzazione dei servizi e il progresso che ne consegue possono aver esito positivo solo se gli utenti si fidano delle tecnologie, del loro utilizzo corretto e della loro sicurezza.

Infatti, maggiore è la fiducia nell'ambiente digitale, più diffuso ed efficace è l'utilizzo delle ICT con la realizzazione dei benefici che ne possono derivare.

Coniata nel 1982, l'espressione *cyber space* indica una realtà globale costituita dai dati e da una rete interdipendente di infrastrutture tecnologiche di informazioni, che includono Internet, le telecomunicazioni e i sistemi di computeraggio.

Superata la tradizionale idea della stessa quale zona franca di libertà quasi assoluta in cui l'unico regime di controllo accettato e accettabile è l'autoregolamentazione, si è posto il problema di una regolamentazione in nome della cybersecurity, espressione piuttosto ampia con cui si allude sia alla tutela della confidenzialità delle informazioni scambiate tramite le infrastrutture digitali e da esse processate sia alla protezione delle reti cibernetiche da attacchi informatici.

In questa direzione si sono orientate sia le Nazioni Unite che l'Unione europea. Nello specifico, i Paesi membri dell'Onu hanno recentemente approvato un rapporto redatto dal c.d. OEWG in cui vengono indicate le raccomandazioni da osservare per realizzare e mantenere la *peace and cybersecurity*, ma soprattutto vengono individuate strutture informatiche critiche – ad esempio, quelle mediche – che richiedono sforzi in ottica multilaterale per aumentarne le capacità di rilevazione e difesa contro le minacce cibernetiche.

Da parte sua, l'Ue ha assunto il ruolo di leader nel rendere lo spazio cibernetico un luogo sicuro e affidabile tanto da fondare la propria strategia per il digitale su un nuovo piano di difesa che condurrà a una società digitale verde e resiliente. A tal fine il Consiglio ha adottato conclusioni sulla strategia in materia di sicurezza informatica presentata dalla Commissione e dall'Alto rappresentante per gli affari esteri e la politica di sicurezza a fine 2020, evidenziando l'esigenza di rafforzare le politiche di diplomazia informatica – soprattutto quando riguardino il contrasto ad attacchi a infrastrutture critiche e servizi essenziali – e promuovere la cooperazione con partner internazionali attesa la portata transfrontaliera della maggioranza delle aggressioni.

3. Conclusioni

Dai paragrafi precedenti emerge come la trasformazione verde e quella digitale siano realizzabili solo a condizione di una significativa cooperazione internazionale, intendendo per tale non solo la tradizionale collaborazione tra attori statali e para-statali, ma anche il coinvolgimento del settore privato (Abba L., 2021).

Questo perché il *cyber space* è ontologicamente privo di barriere fisiche e geografiche, sicché la sua regolamentazione e la sua messa in sicurezza richiedono un coordinamento strategico di tutti i soggetti coinvolti. Infatti, a tal proposito sempre più frequentemente

si parla di diplomazia informatica e di geopolitica del digitale, espressioni con cui si allude all'assetto dei nuovi equilibri di potere nello spazio digitale e all'insieme di metodologie di gestione del rischio condivise dalla comunità internazionale (Darnis J-P. et a., 2019).

La cooperazione multilaterale appare, allora, la principale prospettiva per la garanzia della cybersecurity, ma rappresenta anche la maggiore sfida attesa la crisi del multilateralismo nonché per le profonde differenze di interessi e valori di cui sono portatori gli attori del processo di digitalizzazione.

In conclusione, ad avviso di chi scrive, l'auspicio è che la cooperazione multilaterale possa essere oggetto di ulteriori approfondimenti sia in termini di regolamentazione giuridica che in termini di una crescente formazione sul tema nella prospettiva dell'elaborazione di strategie difensive più efficaci e più efficienti.

Riferimenti bibliografici

Benitez D.A., Fava C. (2019), *Sostenibilità: sfida o presupposto?*, Cedam, Padova.

Fabiatti G., Giovannoni E. (2014), *La sfida della sostenibilità integrata*, *Equilibri – Rivista per lo sviluppo sostenibile*, (2/2014), pp. 371-380.

Abba L. (2021), *Carenze attuali e soluzioni future nei meccanismi per la cooperazione digitale*, *Rivista italiana di informatica e diritto*, (1/2021), pp. 141-150.

Darnis J-P., Polito C. (a cura di) (2019), *La geopolitica del digitale*, Edizioni Nuova Cultura, Roma.

Autrice



Angelica Maiorano - angelica.maiorano@uniba.it

Laureata in giurisprudenza nel 2018 presso l'Università degli Studi di Bari "Aldo Moro", nell'ottobre 2019 inizia il Dottorato in "Principi giuridici e istituzioni fra mercati globali e diritti fondamentali" presso lo stesso ateneo, curriculum di diritto internazionale e dell'Unione europea, occupandosi di protezione dei dati personali e mercato unico digitale.

Nel corso dell'a.a. 2020/2021 prende parte a diversi progetti internazionali per lo studio e l'approfondimento di temi di respiro europeo.

Carbon Footprint Italy: how blockchain technology supports innovation in tackling climate change

Daniele Pernigotti

Aequilibria S.r.l. – SB

Abstract. Climate change is the most pressing challenge human beings must face in the present and the near future. In a growing need for increasingly ambitious mitigation actions, Carbon Footprint Italy is the national Programme Operator specifically dedicated to communicating the results of greenhouse gas emissions quantification for both products and organisations. This paper explores the innovative Carbon Footprint Italy experience, showing how blockchain technology helped to enhance the program reliability and transparency.

Keywords. climate change, carbon footprint, programme operator, blockchain, innovation

Introduction

Climate change is broadly recognised as the main challenge human beings will need to cope with in the present and near future. The awareness of this urgency is not only widespread among the scientific community but within all sectors of our society.

On the part of industries, this implies a new deal between producers and consumers in jointly promoting a low carbon market and effectively communicating the efforts undertaken. Consumers and clients, in fact, often face difficulties in recognising reliable products, and therefore risk falling into the so-called greenwashing – either consciously or not.

These challenges led to the development of specific subjects, known as Programme Operators (POs), dedicated to the communication of carbon footprints with the purpose of ensuring the effective transmission of commitments.

1. The importance of quantification

The first step for any greenhouse gas (GHG) emissions mitigation action is a solid and transparent quantification of GHG related to products and organisations. It should be noted that not always companies adopt robust and internationally recognised standards within solid verification systems, undermining their public statements. For this reason, it is notable to observe the establishment of Carbon Footprint Italy (CFI), the national Programme Operator that aims at closing this gap and at demonstrating companies' com-



Fig. 1
Carbon
Footprint
Italy logo

mitment to reducing their GHG emissions effectively.

CFI is the only Programme Operator in Italy, affiliated to the international network Carbon Footprint International, specifically dedicated to communicating the results of the quantification of GHG emissions, known as Carbon Footprint (CF) of products and organisations, their reductions, and the attested Carbon Neutrality claims.

2. CFI's values

Based on third parties' verification statements, CFI has been created to ensure proper and reliable communication of the efforts undertaken to provide low-carbon solutions on the market. To do so, CFI is based on four core values: its main strengths are to be credible, guaranteed, flexible, and effective (CFI, 2021).

2.1. Credibility

CFI exclusively provides for the use of internationally recognised standards for the quantification of GHG emissions or their reduction.

ISO, the International Organisation for Standardisation, is the leading worldwide independent organisation to set up of internationally recognised standards (ISO, 2021). In this case, the two most relevant ISO standards recognised by CFI are ISO 14067, for the Carbon Footprint of Product (CFP), and ISO 14064-1, which refers to the Carbon Footprint of Organisations (CFO).

In addition, for what regards the achievement of Carbon Neutrality and the use of its relative mark, CFI takes as reference the standard PAS 2060:2014, which regulates the demonstration of carbon neutrality, defining requirements for quantification, reduction, and offsetting of GHG emissions in specific areas. Although CFI exclusively recognises ISO standards, the PAS 2060 – issued by the British Standards Institution – is commonly recognised at the international level, up to the future development of ISO 14068.

2.2 Guarantee

All products and organisations registered in CFI report values of GHG emissions, or their reduction, supported by verification statements issued by independent third-party verification bodies, accredited by Accredia – the Italian national accreditation body – or equivalent international accreditation bodies.

2.3 Flexibility

CFI is flexible because companies and products registered in the Programme Operator may employ specific marks to disclose their participation in CFP or CFO. Moreover, CFI presents a Carbon Reduction mark, specifically dedicated to demonstrating the reduction of GHG emissions of products and organisations.

Furthermore, to prove the achievement of Carbon Neutrality, a specific mark is available for organisations or products already registered in CFI that have offset the residual GHG emissions: they can do so by purchasing recognised carbon credits after the implementation of mitigation actions.

2.4 Effectiveness

When it comes to effectiveness, transparency and accountability are essential. Concerning transparency, CFI developed a dedicated page for each product and organisation participating in the Programme. These dedicated pages have several background information to share key data related to every single CF. CFI also introduced the innovative blockchain technology to make companies accountable for their CF claims, thanks to unmodifiable information on the participants in the Programme.

3. The innovative blockchain technology

According to several authors, blockchain technologies enable actors to verify and audit transactions and track them in ledgers (Sabeti et al., 2018). Once the transactions are approved, they can be added as a new block in the blockchain. Smart contracts then increase the credibility of transactions. In this way, values are assigned to the original information, and the system tracks them over time. Blockchain technologies are increasingly helpful to implement climate action.

Considering this need, CFI is the first Carbon Footprint Programme worldwide to use blockchain technologies to guarantee the accurate tracking of GHG emissions and any reductions that can occur due to mitigation actions.

3.1 Overview of the benefits of the blockchain technology applied

There are many benefits related to the implementation of blockchain technology. Firstly, it leads to publicly available unmodifiable high-quality records and information regarding CF data (e.g., t CO₂e, verification statements, and so on), both at product and organisation level.

Secondly, this technology makes unmodifiable CF data available over time: this allows for the accurate tracking of the committed mitigation targets and improves the targets' reliability for the transition towards a low-carbon economy.

The project's development, including the direct integration of data collected in the blockchain, will further increase the transparency of the assessment of GHG emissions, eliminating the risk of fraud and greenwashing.

4. Conclusions

CFI decided to introduce blockchain technology to make companies accountable for their CF claims through unmodifiable information on who participates in the Programme. Blockchain allows for transparent and reliable control over time of the reported information by any stakeholder and addresses progress that may be quickly and easily evaluated. At the same time, the open-access platform can be consulted by everyone, which means that the information can be shared with all relevant stakeholders, including policymakers. Thus, information disclosed by the companies and verified by an accredited third-party body cannot be modified over the years when registered in CFI.

Overall, CFI represents a concrete example of how digital tools can support the adoption of reliable carbon tools that promote the transition towards a low carbon economy.

Bibliography

Carbon Footprint Italy, 'The Programme', <https://carbonfootprintitaly.it/en/the-programme/>, visited on 08.09.2021.

ISO (2021), 'Popular Standards' <https://www.iso.org/popular-standards.html>, visited on 08.09.2021.

Saberi S., Kouhizadeh M., Sarkis J., Shen L. (2018), Blockchain technology and its relationships to sustainable supply chain management, *International Journal of Production Research* (57:7), pp. 2117-2135.

Author



Daniele Pernigotti dpernigotti@aequilibria.com

Daniele Pernigotti is the Director of Carbon Footprint Italy. He is in addition the CEO of Aequilibria S.r.l – SB, Chair of CEN/TC 467 and of the Italian mirror group of ISO/TC 207 on climate change. He also coordinated the development of the standard ISO 14067:2018 on the Carbon Footprint of Product (CFP), and coordinates the ISO/TC 207 TG2 on Circular Economy. He is external technical expert on GHG for Accredia and is also qualified as Lead Assessor on GHG for ANAB (USA) and ONAC (Colombia).

An Efficient and Privacy-Aware Method for Revealing Network Covert Channels

Marco Zuppelli, Luca Caviglione, Corrado Pizzi, Matteo Repetto

IMATI - Institute for Applied Mathematics and Information Technologies

Abstract. The ability of creating covert channels within network traffic is increasingly exploited by malware to elude detection and remain unnoticed. Unfortunately, spotting such hidden communication attempts often requires to evaluate composite and huge volumes of data, which may lead to scalability and privacy issues. We propose an efficient method for computing suitable indicators to reveal the presence of covert channels within the bulk of traffic. To meet performance criteria, we exploit code augmentation features of the Linux kernel. Privacy is guaranteed by using a counter-based mechanism not requiring to store information of the header. To prove its effectiveness, we tested our idea with a covert channel targeting IPv6 traffic.

Keywords. Covert channels, code augmentation, stegomalware, detection.

Introduction

Attacks like ransomware, cryptominers, and advanced persistent threats daily endanger individuals and large-scale organizations, highlighting the limits of common security tools and causing relevant economical losses. Among the various offensive techniques, an emerging trend concerns the adoption of information hiding or steganography to create attacks difficult to spot or detect. Accordingly, such a group of threats has been named stegomalware. In general, stegomalware implements methodologies to hide malicious routines or configuration files within innocent-looking pictures, bypass firewalls, implement stealthy multi-stage loading architectures, and force execution enclaves provided by sandboxing and virtualization (Caviglione et al. 2020).

A key component for the success of stegomalware is rooted in its ability to create network covert channels. In essence, a network covert channel allows two remote endpoints to secretly exchange data via the injection of arbitrary information within legitimate traffic flows (Zander et al. 2007). Unfortunately, detecting or mitigating such abusive communications pose several challenges. First, the feature exploited by the attacker is not known a priori, thus requiring to inspect various traffic entities (e.g., header fields, payloads or behaviors like the throughput or the inter-packet delay) at the same time. Second, adopting deep inspection techniques may cause scalability and performance issues. Third, gathering traffic information introduces additional fragilities in the privacy of users. In fact, even in the presence of suitable anonymization techniques, large volumes of data could make statistical guessing attacks a real concern (Burkhart et al. 2010).

Therefore, this chapter introduces a framework exploiting code augmentation features

of the Linux kernel to efficiently inspect traffic and compute privacy-preserving metrics that can be used to reveal the presence of covert communications. The contributions of this chapter are:

- knowledge transfer: the diffusion of information-hiding-capable threats requires their understanding to completely assess the security of modern digital infrastructures ;
- support the development of digital infrastructures: enforcing security while guaranteeing suitable performance metrics is relevant to develop next-generation services. Yet, the use of lightweight mechanisms can mitigate hardware costs and the complexity of the software needed to gather information;
- data usage: the Internet contains data that can be used to profile or track users. Thus, being able to detect exfiltration attempts is of prime importance, while the development of privacy-by-design detection mechanisms or countermeasures can encourage their deployment.

The rest of the chapter is structured as follows. Section 2 showcases how network covert channels can be detected by using indicators not leaking or collecting sensitive information, while Section 3 concludes the chapter and portrays possible future research directions.

2. Covert Channel Detection via Counters

To prove the effectiveness of our idea, in this chapter we consider attacks exploiting covert channels targeting the header of the IPv6 protocol, which have been observed in real threats or are expected to become a concern in the near future (Zander et al. 2007, Caviglione et al. 2021). As a paradigmatic example, in the following we discuss network covert channels built via the embedding of secret data in the Flow Label field.

The typical approach for detecting covert communications requires to trace network traffic with a per-flow granularity. For each stream, parameters like the number of packets, data transmitted on the wire and the average inter-packet delay are recorded. Values of fields suspected to contain secrets have to be stored, too. As a result, the memory footprint could be prohibitive and recorded information can disclose personal details or the user identity. To cope with such drawbacks, we propose a technique able to scale independently of the number of flows. Rather than keeping the “state” for each flow, we only count the number of occurrences for the different values that a given field assumes. To enforce scalability, multiple values may be grouped together into a “bin”, and a single counter is used for the whole group. For instance, for the case of the Flow Label, bins can be a partition of its 20-bit space leading to $B \leq 2^{20}$ equally-sized containers. This approach, named as counters, guarantees privacy and anonymity, since no values of the field under investigation are stored. To pursue efficiency, the data collection phase has been implemented via the extended Berkeley Packet Filter (eBPF), which introduces a lightweight overhead and allows to have a general framework to gain visibility over network and software.

To prove the effectiveness of the approach, we present results partially borrowed from our ongoing research (Caviglione et al. 2021). Since each IPv6 conversation has a unique Flow Label, the number of bins N observed in a time window T can provide a rough estimation of the number of active flows. Discrepancies between a known behavior or measurements of an external monitoring tool (denoted as F in the following) can spot an

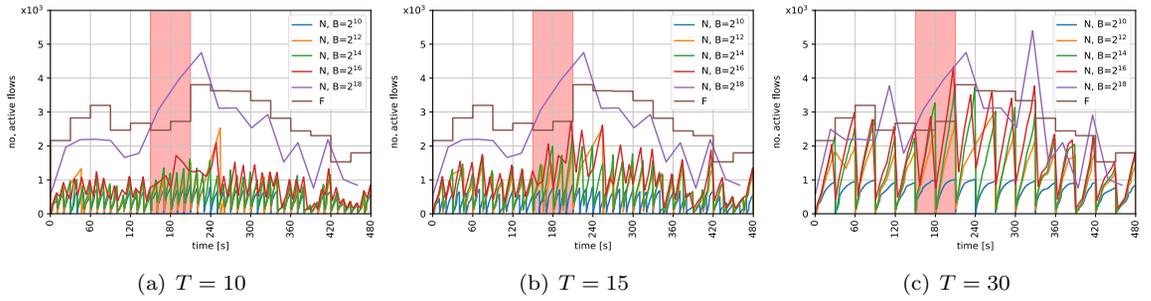


Fig. 1

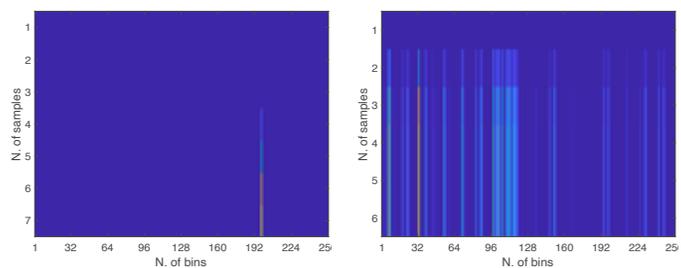
Number of changing bins of the inspected traffic for different observation windows of T seconds. Red areas denote when a covert communication is present

attacker maliciously altering the field to exfiltrate data or to orchestrate an attack. Figure 1 shows the idea.

The number of changing bins can be also organized in a heatmap to produce a synthetic image capturing the evolution of the various values assumed by the Flow Label. Specifically, Figure 2(a) represents a “clean” network conversation, i.e., only the bin related to the licit Flow Label value increases, while Figures 2(b), 2(c), and 2(d) show a flow containing a covert channel exfiltrating text, a JPG image or random data, respectively. This “pictorial” representation of the channel does not contain any sensitive information and can feed AI-based frameworks to detect threats and classify the exfiltrated data (e.g., a .dll or a shellcode).

3. Conclusions and Future Works

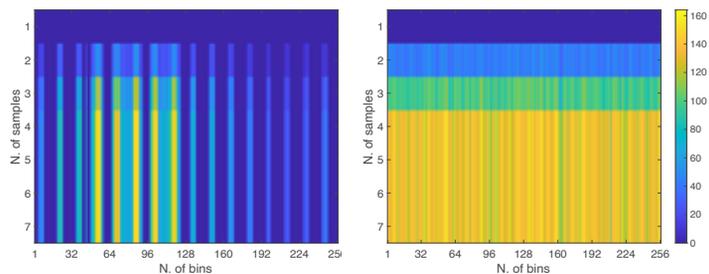
This chapter presented a technique leveraging code augmentation for capturing the behavior of traffic in an efficient and privacy-preserving manner. The use of eBPF allows to easily extend the collection phase to other fields, protocols or traffic behaviors, thus improving the extensibility of the approach. Future works aim at refining the idea, especially



(a) Clean

(b) Text

Fig. 2
Heatmaps for various covert transmissions, with $B=28$. Heatmaps have been computed only considering packets belonging to the covert channel



(c) JPG

(d) Random

to identify the exfiltrated content or the class of malware using the covert channel. Part of the ongoing research is aimed at integrating the eBPF-based detection mechanism within cloud-based architectures or novel security toolkits.

References

Burkhart M., Schatzmann D., Trammell B., Boschi E., Plattner, B. (2010), The role of network trace anonymization under attack, ACM SIGCOMM Computer Communication Review, 40(1), pp. 5-11.

Caviglione L., Mazurczyk W., Repetto M., Schaffhauser A., Zuppelli M. (2021). Kernel-level tracing for detecting stegomalware and covert channels in Linux environments, Computer Networks, (191), pp. 1-12.

Caviglione L., Choraś M., Corona I., Janicki A., Mazurczyk W., Pawlicki M., Wasielewska K. (2020), Tight arms race: overview of current malware threats and trends in their detection, IEEE Access, (9), pp. 5371-5396.

Zander S., Armitage G., Branch P. (2007), A survey of covert channels and countermeasures in computer network protocols, IEEE Communications Surveys & Tutorials, 9(3), pp. 44-57.

Authors



Marco Zuppelli marco.zuppelli@ge.imati.cnr.it

Marco Zuppelli is a research fellow at IMATI-CNR. He investigates novel detection methods for steganographic malware exploiting both network and local covert channels.

Luca Caviglione luca.caviglione@ge.imati.cnr.it

Luca Caviglione is a Senior Research Scientist at IMATI-CNR. Its prime research topics are information hiding, network covert channels and networking.



Corrado Pizzi corrado.pizzi@ge.imati.cnr.it

Corrado Pizzi is a technologist at IMATI-CNR and he graduated in Mathematics at Genoa University. He studied techniques of geometric reasoning for the extraction of shape features from discrete surface models and methods for data set simplification. He is responsible of the IMATI 3D scanning and printing laboratory and of the IMATI information systems and network services. His interests are also in network security and in system management.

Matteo Repetto matteo.repetto@ge.imati.cnr.it

Matteo Repetto is a Research Scientist at IMATI-CNR and he is the Scientific and Technical Coordinator of the EU projects ASTRID and GUARD. His research interests include network functions virtualization and service functions chaining, network and service security.



Riconoscimento vocale di formule

Matteo Amore⁴, Tiziana Armano¹, Cristian Bernareggi³, Anna Capietto¹, Sandro Coriasco¹, Roberta Crespan⁴, Mattia Ducci¹, Maria Luisa Gabrielli⁴, Maria Francesca Guadalupi⁴, Alessandro Mazzei², Antonio Mazzei⁴, Adriano Sofia¹, Francesco Tarasconi⁴

¹Dipartimento di Matematica – Università di Torino, ²Dipartimento di Informatica – Università di Torino, ³Università di Milano, ⁴Maize srl

Abstract. Le tecnologie in generale, ed in particolare le tecnologie assistive e gli strumenti compensativi, hanno attualmente un livello di sviluppo elevato ma tuttavia presentano criticità nella gestione di contenuti scientifici (formule e grafici). Queste problematiche disincentivano persone con disabilità e disturbi specifici di apprendimento (DSA) ad accedere a percorsi di formazione di tipo scientifico e precludono in seguito la possibilità di accedere a professioni STEM. Il Laboratorio per la Ricerca e la Sperimentazione di Nuove Tecnologie Assistive per le STEM "S. Polin" del Dipartimento di Matematica "G. Peano" dell'Università di Torino si occupa di sviluppare nuove soluzioni tecnologiche e testare, adattare e diffondere le tecnologie già esistenti allo scopo di risolvere i problemi relativi all'accesso dei contenuti STEM (Armano et al, 2020). In questo ambito il Laboratorio si occupa anche di ricerca e sviluppo in relazione al riconoscimento vocale di formule.

Keywords. Accessibilità, formule, riconoscimento vocale, disabilità, DSA.

Introduzione

Il progetto Matematica a voce, cofinanziato da Fondazione CRT, nasce dall'esigenza di avere sistemi per la dettatura e la trascrizione di formule che sarebbero molto utili a persone con disabilità motorie (permanenti o temporanee), uditive e a persone con DSA.

L'“oggetto” video in generale ed in particolare i video di lezioni (molto diffusi in questo periodo in cui prevale la didattica a distanza) presentano diversi problemi di accessibilità: servono sottotitoli adeguati per persone con disabilità uditive e la trascrizione per persone con disabilità motorie, visive e con DSA che hanno difficoltà a prendere appunti. Esistono attualmente numerosi sistemi per la trascrizione e la sottotitolazione automatica ma nessuno gestisce le formule in modo opportuno: per la fruizione di un video con contenuti scientifici serve la trascrizione di formule in linguaggio specifico (LaTeX o MathML).

I sistemi di interazione vocale, molto diffusi ed evoluti, hanno scarse funzionalità di gestione del linguaggio matematico mentre le persone con disabilità motorie agli arti superiori, permanenti o temporanee, avrebbero bisogno di strumenti per scrivere ed elaborare in autonomia le espressioni matematiche. Alcuni software promettenti in questo campo come MathTalk ed Equatio funzionano solo per la lingua inglese.

Il nostro progetto prevede due moduli relativi a trascrizione e sottotitolazione di video di lezioni con formule e dettatura e gestione di formule matematica:

1. VoiceMath: sviluppo di software per la trascrizione in testo + LaTeX (o MathML) di lezioni universitarie “con formule”, utile a persone con disabilità motorie, uditive e con DSA;
2. SpeechMatE: sviluppo di software per la dettatura e la modifica di formule matematiche con il riconoscimento della lingua italiana, utile a persone con disabilità motorie.

1. VoiceMath

La soluzione ha come obiettivo la produzione automatica di un flusso ordinato di contenuti elettronici (txt e LaTeX) a partire da una videoregistrazione di lezioni e corrispondenti immagini del testo scritto a mano dall'insegnante. Le materie in oggetto sono: algebra/logica, analisi, fisica matematica, geometria e probabilità/statistica. Per raggiungere questo scopo saranno sviluppate componenti basate su Machine Learning, in particolare di Speech-to-Text e Computer Vision. Partendo dalla videoregistrazione di una lezione reale bisognerà analizzare, con strumenti e modalità specifiche, sia la componente audio, cioè il parlato del docente, sia le immagini, costituite dal testo scritto dal docente durante la lezione.

L'audio, isolato dalla videoregistrazione, dovrà essere sottoposto a trascrizione automatica mediante tecnologia di Natural Language Processing Speech-to-Text (Jurafsky, Manning, 2012). Attualmente è in uso il trascrittore Nuance Transcription Engine (NTE3). Il progetto ha avuto una fase esplorativa, con obiettivo quello di verificare eventuali problematiche degli audio, valutare l'accuratezza della trascrizione automatica ed effettuare le dovute attività di post processing. Per effettuare la valutazione dell'accuratezza e gestire le miglitorie, siamo partiti da dalle trascrizioni manuali delle lezioni (Savy 2005), che sono valse come standard. Il focus di queste lezioni, ovvero le formule, è stata la parte più sfidante da dover trascrivere: i trascrittori automatici sono generalmente calibrati attorno a modelli linguistici generalistici, mentre i nostri audio trattano una terminologia specifica. In linguistica questa viene definita come terminologia di dominio. Per migliorare la trascrizione anche di formule ed espressioni specifiche del linguaggio matematico abbiamo prodotto un file specifico detto DLM (Domain Language Model), contenente tali termini, per un numero ridotto di audio; tale file ha come obiettivo quello di andare ad addestrare il trascrittore, per renderlo più tarato su quello che è il dominio applicativo. Il test effettuato ha evidenziato delle miglitorie per le trascrizioni automatiche, motivo per cui puntiamo ad ampliare tale DLM e ad utilizzarne uno specifico per insegnamento. Al momento, siamo ancora in una fase di sperimentazione e di analisi prevalentemente qualitativa degli errori, quindi non escludiamo di portare avanti nuovi test successivamente.

Il modulo “Image to Latex” si occuperà di identificare in modo automatico le formule contenute in un'immagine e di convertirle in una stringa LaTeX. Il dataset finora conta circa 50 immagini per ognuna delle materie analizzate. In una prima fase ci siamo occupati di analizzare le performance di Mathpix, tool che si occupa della “digitalizzazione” di testi stampati o handwritten con contenuti matematici e scientifici in generale. Le immagini originali, corrispondenti solitamente a intere pagine, generalmente contengono molte in-

formazioni: per migliorare le analisi delle componenti successive, le abbiamo segmentate isolando le parti significative contenenti le formule. È stata posta molta attenzione alla preparazione di un dataset rappresentativo, che contenesse un'ampia gamma di simboli e terminologia del dominio matematico. Terminata questa attività di preprocessing, le immagini sono state convertite in LaTeX mediante tecnologie di Computer Vision. Anche per questo modulo, stiamo effettuando delle analisi qualitative degli errori per capire in generale le performance del tool ed identificare eventuali problematiche sul riconoscimento di testi handwritten o di particolari simboli/caratteri, che permetteranno di decidere i nuovi test da portare avanti.

Al termine di queste attività di Speech-to-Text e Computer Vision, gli output ottenuti dovranno confluire in un unico flusso, che consentirà agli studenti di ripercorrere ordinatamente la lezione. Alle formule LaTeX viene associato un demarcatore temporale che consente di inserirle al punto opportuno della lezione.

2. SpeechMatE

Il riconoscimento vocale ha avuto negli ultimi anni un significativo progresso: sono disponibili applicazioni azionabili tramite comandi vocali, assistenti personali intelligenti basati su interfacce di dialogo, sistemi operativi con assistenti vocali. Tuttavia ci sono pochi tentativi di sviluppare soluzioni per il riconoscimento di formule matematiche, soprattutto in lingua italiana, anche se l'utilità di tali soluzioni sarebbe di grande impatto per un'alta percentuale della popolazione. Sicuramente questo è dovuto ad alcune difficoltà tra cui: la rappresentazione non lineare delle formule, l'ambiguità della lettura di formule matematiche e l'impatto di percentuali di errore nel riconoscimento.

In questo ambito il Laboratorio ha sviluppato il prototipo del software SpeechMatE, per la dettatura e la modifica di formule con riconoscimento della lingua italiana. Il software ha cinque principali componenti: speech recognizer, parser, LaTeX editor, LaTeX compiler, e PDF viewer. L'interazione con il software avviene con comandi vocali, dettatura di testo o formule. L'input vocale viene processato dallo speech recognizer che lo trascrive in lingua italiana. Il testo viene poi processato dal parser che converte la parte di matematica in un frammento di codice LaTeX, che viene successivamente passato all'editor LaTeX e trasformato in PDF. Il modello di interazione vocale utilizza tre ambienti attivati dalle relative parole chiave: testo, matematica e comandi (per i comandi di edit). Dopo una fase di confronto sono stati scelti, per il riconoscimento vocale, i servizi di Google Cloud e, come editor LaTeX, TeXStudio. Attualmente viene gestita una parte limitata di simboli e espressioni matematiche: l'estensione di queste funzionalità è attualmente in corso di sviluppo (Bernareggi et al, 2020).

Riferimenti bibliografici

Armano T., Capietto A., Ahmetovic D., Bernareggi C., Coriasco S., Ducci M., Magosso C., Mazzei A., Murru N., Sofia A. (2020), Accessibilità di contenuti digitali per le STEM: un problema aperto. Alcune soluzioni inclusive per l'accessibilità di formule e grafici, Mondo Digitale, vol. 89, pag. 1-12.

Bernareggi C., Ducci M., Mazzei A., Ahmetovic D., Armano T., Capietto A., Coriasco S., Murru N. (2020), SpeechMatE: A Speech-driven Maths Editor for Motor-Impaired People, ICCHP open access compendium "Future Perspectives of AT, eAccessibility and eInclusion".

Jurafsky, D., & Manning, C. (2012). Natural language processing. *Instructor*, 212(998), 3482.

Savy, R. (2005), "Specifiche per la trascrizione ortografica annotata dei testi" in *Italiano Parlato. Analisi di un dialogo*, Liguori, Napoli.

Autori

Amore Matteo matteo.amore@h-farm.com

Laureato in Linguistica presso l'Università di Pavia, tesi sulla semantica distribuzionale. È un linguista computazionale che si occupa di attività di estrazione di testo per progetti di Customer Experience, ossia Text Classification, Sentiment Analysis, Opinion Mining, Named Entities Recognition. Si occupa inoltre di: progettazione e valutazione di corpora di training; test di sistemi di Automatic Speech Recognition, valutazione di tecnologie vocali, annotazione di dati per progetti di ricerca.

Tiziana Armano tiziana.armano@unito.it

Tecnico della ricerca presso il Dipartimento di Matematica "G. Peano" dell'Università degli Studi di Torino. Si occupa di e-learning e tecnologie per la didattica. Ha fatto parte del gruppo di ricerca del progetto "Per una matematica accessibile e inclusiva" dal 2012 ed è membro del Laboratorio Polin. In questo ambito ha seguito in particolare lo sviluppo del pacchetto Axessibility e attualmente si occupa del progetto Matematica a voce.

Cristian Bernareggi cristian.bernareggi@unimi.it

Dottorato di ricerca in informatica nel 2006. Tra il 2004 e il 2009 ha condotto ricerca in ambito HCI nei progetti europei LAMBDA ed @Science. Dal 2007 è collaboratore tecnico scientifico in ambito informatico per l'Università di Milano. Tra il 2011 e il 2018 ha partecipato alla fondazione e alla crescita dello spin-off EveryWare Technologies per lo sviluppo di tecnologie assistive su dispositivi mobili. Dal 2018 collabora con il Laboratorio Polin sull'accessibilità alle STEM per persone con disabilità.

Anna Capietto anna.capietto@unito.it

Professore Ordinario di Analisi Matematica. Da 15 anni è Referente per le tematiche degli studenti con disabilità presso il Dipartimento di Matematica. Dal 2013 è responsabile del progetto "Per una matematica accessibile e inclusiva", le cui attività hanno condotto nel novembre 2018 all'istituzione del Laboratorio per la ricerca e la sperimentazione di nuove tecnologie assistive per le STEM "S.Polin". È presidente del Comitato Public Engagement dell'Università di Torino.

Sandro Coriasco sandro.coriasco@unito.it

Professore Associato di Analisi Matematica presso il Dipartimento di Matematica dell'Università di Torino. Si occupa di analisi microlocale e di analisi globale su varietà non compatte e su varietà con

singularità. È un componente del gruppo di ricerca sulle tecnologie assistive presso il Laboratorio "S. Polin". In tale ambito si occupa, in particolare, dello sviluppo del pacchetto LaTeX accessibility e, più in generale, dei temi collegati all'accessibilità dei contenuti matematici.

Roberta Crespan roberta.crespan@h-farm.com

Laureata in Mediazione Linguistica e Culturale presso l'Università di Padova. Ha approfondito le conoscenze nell'ambito della Traduzione audiovisiva e Accessibilità Audiovisiva. Dal 2019 lavora come linguista presso CELI Language Technology dove ha avuto modo di avvicinarsi al mondo della linguistica computazionale legata all'NLP. Svolge attività di test sui chatbot, di verifica dei flussi. Ha partecipato ad attività legate a progetti di ASR e si occupa di localizzazione e UX Writing di contenuti.

Mattia Ducci mattia.ducci@unito.it

Laurea magistrale in informatica all'Università degli Studi di Milano nel 2019. Durante e dopo l'università ha collaborato a vari progetti incentrati sul tema dell'accessibilità: da soluzioni su dispositivi mobili per persone con disabilità visive (Università degli Studi di Milano) a soluzioni per rendere più accessibili le materie scientifiche (Università degli Studi di Torino). Dal 2018 lavora come consulente informatico in un'azienda per sviluppo mobile iOS e cross-platform e full-stack.

Maria Luisa Gabrielli marialuisa.gabrielli@h-farm.com

Lavora come Data Scientist presso Celi Language Technology. Laureata in Matematica presso l'Università di Torino, si è specializzata in Analisi numerica ed in particolare modo si è avvicinata al mondo dell'elaborazione e del riconoscimento di immagini. Ha esperienza in Data Analysis, nell'applicazione di algoritmi di Machine Learning in ambito Forecasting e in tecniche di Computer Vision.

Mariafrancesca Guadalupi mariafrancesca.guadalupi@h-farm.com

Lavora come linguista computazionale dopo aver conseguito una laurea magistrale in Linguistica e Comunicazione Persuasiva, Tecnologia e Studi Cognitivi ed un Master di II Livello in Intelligenza Artificiale. Si occupa di progetti riguardanti: assistenti conversazionali, sentiment analysis ed opinion mining, text classification. Sta studiando per approfondire il mondo delle tecnologie assistive.

Alessandro Mazzei alessandro.mazzei@unito.it

Laurea in Fisica (indirizzo Cibernetico) presso l'Università Federico II di Napoli nel 2000. Dottore di Ricerca in Informatica presso l'Università degli Studi Torino nel 2005, dove è ricercatore presso il Dipartimento di Informatica. I suoi interessi di ricerca includono le ontologie giuridiche, la sintassi e la semantica computazionale, la generazione automatica del linguaggio, i sistemi di dialogo, il ragionamento automatico, le tecnologie assistive e il food computing. È autore di oltre 100 lavori scientifici peer reviewed.

Antonio Mazzei antonio.mazzei@h-farm.com

Si è laureato al Politecnico di Torino ed all'ENST di Parigi in ingegneria delle telecomunicazioni. Ha rivestito molti dei ruoli IT fino all'attuale posizione da Senior Manager. La sensibilità per le

peculiarità di ciascun ruolo, lo spingono a bilanciare le esigenze del team e valorizzarne le capacità al fine di raggiungere gli obiettivi di progetto.

Specializzato in sistemi di Enterprise Search ed integrazione di sistemi di NLP ed AI nei processi aziendali.

Adriano Sofia adriano.sofia@unito.it

Assegnista di ricerca presso il Laboratorio "S. Polin" dell'Università degli Studi di Torino dove si occupa di sonificazione e accessibilità. Conseguita la Laurea Magistrale in Fisica con Lode i suoi interessi scientifici riguardano due filoni: il primo incentrato sulle simulazioni di dinamica molecolare classica ed ab initio applicata alla fisica della materia e alla biochimica, e il secondo riguarda le tecnologie assistive e l'accessibilità.

Francesco Tarasconi francesco.tarasconi.h-farm.com

É Senior Data Scientist presso CELI Language Technology, ora centro di Intelligenza Artificiale di H-Farm Innovation. Laureato in Matematica presso l'Università di Torino, ha condotto un progetto di ricerca applicata sui Sistemi Complessi con il supporto di CELI e Fondazione CRT. Lavora da quasi dieci anni in progetti di Data Science, sia in ambito business che di ricerca; si interessa principalmente a Unsupervised Machine Learning, Forecasting e Deep Language Models.

CHNet cloud: an EOSC-based cloud for physical technologies applied to cultural heritages

Alessandro Bombini¹, Lisa Castelli¹, Luca dell'Agnello², Achille Felicetti³,
Francesco Giacomini², Franco Niccolucci³, Francesco Taccetti¹

¹Cultural Heritage Network, Istituto Nazionale di Fisica Nucleare, Florence. ²CNAF, Istituto Nazionale di Fisica Nucleare, Bologna. ³VAST-LAB, PIN, University of Florence

Abstract. In the framework of the European projects ARIADNEplus and EOSC-Pillar, a cloud system, named Tools for HEritage Science Processing, Integration, and ANalysis (THESPIAN), was developed, which offers multiple microservices to the researchers of the Cultural Heritage Network (CHNet) of INFN (Istituto Nazionale di Fisica Nucleare), from storing their raw data to reuse them by following the FAIR principles for establishing integration and interoperability among shared information. The CHNet cloud currently offers three web services: THESPIAN-Mask, a service for assisted metadata generation and data storage, based on an ad-hoc developed ontology called CRMhs; THESPIAN-NER, a tool based on a deep neural network for Named Entity Recognition, which can interpret Italian-written archeological documents and annotate them extracting named entities, that are used for devising custom CHNet database queries; and XRF analyser, a tool for on-line, real-time elaboration of raw data of X-Ray Fluorescence imaging analysis performed on Cultural Heritage.

Keywords. cloud services, FAIR Data management, Big data and Smart Data, AI Machine and Deep Learning.

Introduction

In the framework of the European projects ARIADNEplus and EOSC-Pillar initiative of the European Open Science Cloud (EOSC) framework, Tools for HEritage Science Processing, Integration, and ANalysis (THESPIAN) was developed. THESPIAN is a cloud system offering multiple microservices to the researchers of the internal INFN network devoted to the application of physical technologies to Cultural Heritage, the so-called Cultural Heritage Network (CHNet), and to all the external researchers who cooperate with the network. The mission of CHNet is to harmonise and enhance the expertise of the Institute in the field of Cultural Heritage, expertise distributed among many structures spread over the whole Italian territory. CHNet includes the INFN research groups whose activity is devoted to the development and application of technologies for the study and conservation of Cultural Heritage, but it is open also to other national partners with expertise complementary to that of the Institute and to international partners engaged in diagnostics in cultural heritage. At the moment, the network is composed of three layers:

1. first level nodes: INFN structures;
2. second level nodes: Partners with complementary competencies, such as restoration

centres or groups of researchers in universities;

3. third level nodes: International partners, such as universities or research institutions from abroad.

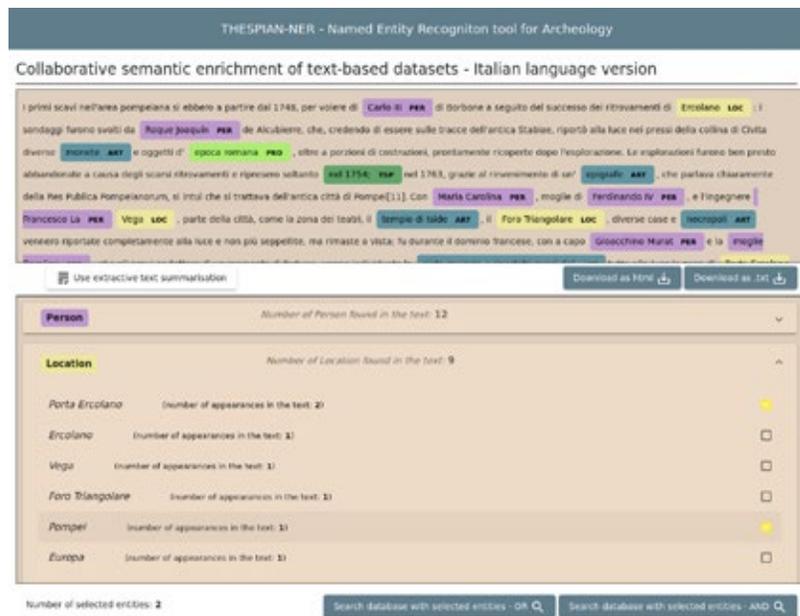
The purpose of the platform is to create a complete ecosystem for scientific data and metadata of physical analysis on cultural heritage, modeled according to the latest ontologies and standards, and to make them interoperable with data generated by other disciplines and accessible on other platforms according to the FAIR principles for data: findability, accessibility, interoperability, and reusability [Wilkinson et al. (2016)].

The Cloud services

Firstly, a web tool dubbed Tools for HEritage Science Processing, Integration and Analysis -Mask (THESPIAN-Mask) has been developed by the Digital Heritage Laboratory (DHLab) of CHNet, and by VAST-LAB, PIN. It consists of a web platform for assisted metadata generation and a service for persistent storage of scientific data and their metadata. The tool is based on CRMhs, an extension of the CIDOC CRM [CRM (v. 7.1.1)] ontology designed for modeling the complex entities typical of heritage science, developed by Istituto Nazionale di Fisica Nucleare and VAST-LAB PIN [Castelli et al. (2019)].

The goal of THESPIAN-Mask service is to offer a cloud environment possessing a multitude of web applications for the storage of (meta)data and their use, in light of the FAIR principles; in particular, the idea is to allow researchers to store their raw data, processed data, results, documentation and article preprints, together on the cloud, with a set of metadata based on a common, shared ontology. The system also allows data and metadata to be accessed by other researchers of the network and (re)used via a set of commonly shared web services, offered entirely on the cloud. The CHNet DHLab cloud thus constitutes a cloud for cooperative data storage and data elaboration of scientific analysis on Cultural

Fig. 1
THESPIAN-NER AI enhanced query service. It is possible to use automatically recognised Named Entities to build a custom query to fetch the server for similar entries.



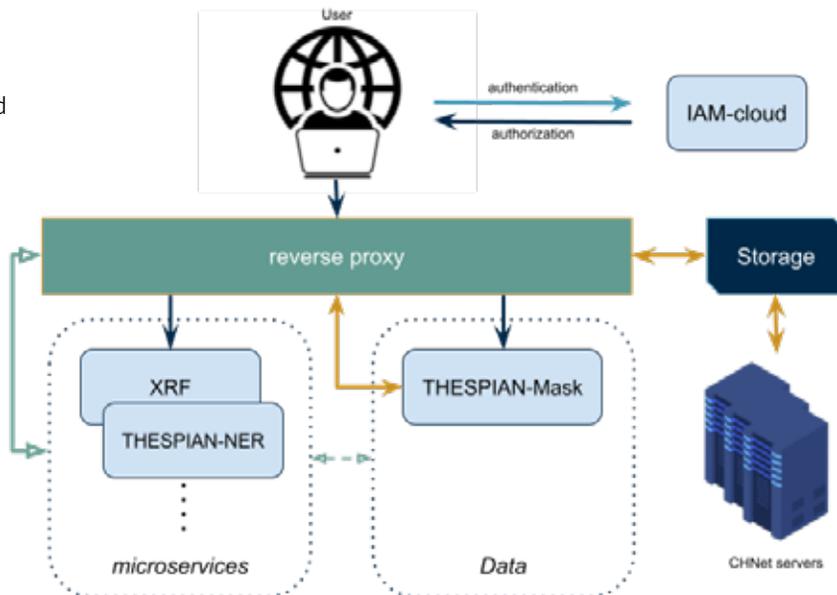
Heritage. The use of CRMhs also makes the information fully interoperable with other CIDOC CRM compatible data and allows their integration in existing cloud environments and extended semantic graphs, such as the semantic data cloud developed by ARIADNE-plus for archaeological data.

Secondly, a web tool called THESPIAN-NER, a Natural Language Processing (NLP) tool for automatic Named Entity Recognition (NER) using deep learning techniques (namely a Convolutional Neural Network), was developed. It allows users to automatically annotate archeological documents written in Italian (either .txt or .pdf files) by identifying and labelling relevant semantic entities in the text, extract and define new metadata for the annotated documents out of them, and using them for building custom queries to fetch related records available on the CHNet database.

The Neural network was built using the python package spaCy (<https://spacy.io/>) and trained using transfer learning on 92 Italian written documents containing 5230 entities, annotated using the open-source program INCEpTION (<https://inception-project.github.io/>); the 9 entities to be identified within the text have been chosen for their relevance in the archeological sector and according to their compatibility with top-level classes of CRMhs and the CIDOC CRM ecosystem. They are: Artefact, Site, Person, Timespan, Activity, Organisation, Location, Period, Biological remains.

THESPIAN-NER is available for researchers as a web tool hosted by the CHNet server in the CHNet cloud environment. Since it enables researchers to either annotate locally stored as well as remotely available documents, provided by the CHNet servers through the cloud platform, it constitutes a (micro)service aiming at fostering data (re)usability. Finally, a tool for analysing raw data of XRF imaging is available. It allows users to fetch the XRF raw data in HDF5 format from the server, or to import their own, and retrieve both the histogram of the XRF counts and the XRF Image, shown in interactive plots that

Fig. 2
Architecture
of CHNet cloud



allow users to easily interact with the data representation and modify it, by adjusting all the various parameters. The raw data file content is a dictionary with key “img” and value the rank-3 tensor of shape (height, width, channel depth) representing the image.

The platforms hosting the cloud have a modular architecture based on containers: the first one hosts a web server offering a graphical web application to input and search data and institutions; the second one is a server processing requests from the web application and interacting with the database deployed within a third container. In front of these a fourth container hosts a reverse proxy that acts as an SSL/TLS terminator and enforces access control policies through a token-based authorisation mechanism relying on the IAM service [Ceccanti et al. (2019)] developed by the INDIGO project [INDIGO-DataCloud Collaboration (2017)]. The container architecture allows easy deployment of the whole cloud service. Also, the front-end part was developed using Angular with TypeScript, while the various RESTful APIs are either written using Node.js and/or Django; the latter was chosen to easily embed the deep neural network developed with Python. To persist data storage, the NoSQL database MongoDB is employed.

Conclusions

The CHNet cloud is the result of the collective efforts of PIN and CHNet: CHNet provided the hardware set-up and developed the web services; CHNet researchers are currently testing the services and populating the database with their data. PIN has contributed by developing the CRMhs ontology on which the metadata are based, by preparing the training dataset of annotated documents for THESPIAN-NER, and also by testing and populating the database; the CHNet-CNAF node has supervised the web development and has devised the cloud architecture, developing also the IAM-cloud authentication service based on OAuth2 [Hardt (2012)].

The digital infrastructure, with all its services, is currently in the pre-production stage, where only a few researchers can access the system, mainly to test the digital infrastructure and web services, ingest preliminary data and check if the design of the system fits their needs. The system will be open to all the researchers of the network for further general tests and to ingest generic scientific data on cultural heritage; the orders of magnitude the system will handle are about $O(100)$ data report per network’s node inserted in the beta testing phase, with $O(10)$ nodes in the network, each of it consisting of $O(10)$ researchers.

Acknowledgments

The present research has been partially funded by the European Commission within the Framework Programme Horizon 2020 with the projects ARIADNEplus (GA no. H2020-INFRAIA-01-2018-2019-823914) and EOSC-Pillar (GA no. H2020-INFRAEOSC-05-2018-2019-857650)

References

Wilkinson, M. et al. “The FAIR Guiding Principles for scientific data management and stewardship”. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

"CIDOC CRM. International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM). Version 7.1.1. <http://www.cidoc-crm.org/version/version-7.1.1>

Castelli, L., Felicetti, A., Proietti, F.: "Heritage Science and Cultural Heritage: a CIDOC-CRM-enabled model for Integration and Interoperability", International Journal on Digital Libraries, Special issue on FAIR Data and Cultural Heritage data-centric research, 2019. DOI: <https://doi.org/10.1007/s00799-019-00275-2>

Hardt, Dick. "RFC6749 - The OAuth 2.0 Authorization Framework". Internet Engineering Task Force, (October 2012). <https://tools.ietf.org/html/rfc6749>

Ceccanti A., Vianello E., Caberletti M.i and Giacomini F., "Beyond X.509: token-based authentication and authorization for HEP", EPJ Web Conf., 214 (2019) 09002, <https://doi.org/10.1051/epjconf/201921409002>

INDIGO-DataCloud Collaboration, INDIGO-DataCloud: A data and computing platform to facilitate seamless access to e-infrastructures, CoRR abs/1711.01981, (2017), arXiv:1711.01981

Authors



Alessandro Bombini bombini@fi.infn.it

Technological Researcher at INFN-CHNet, Florence section, is responsible for the research and development of digital infrastructure and web services for the Digital Heritage Laboratory (DHLab), within the European projects European Open Science Cloud (EOSC) - Pillar (<https://www.eosc-pillar.eu/>), and ARIADNEplus (<https://ariadne-infrastructure.eu/>).

Lisa Castelli castelli@fi.infn.it

Graduated in Physics, has obtained her Ph.D. in Science and Engineering of Materials, and is a Technologist at INFN since June 2019; Lisa Castelli is an expert in the development and applications of portable X-ray-based instrumentation for the elemental characterisation of materials in cultural heritage. She is also involved in the management and organization of the activities of the INFN-CHNet network, and she actively works on the integration of the CHNet raw data, mainly in the framework of the European Projects ARIADNEplus and EOSC - Pillar.



Luca dell'Agnello luca.dellagnello@cnafe.infn.it

Director of Technology at CNAF, INFN. He was a Network expert for GARR network (1997-2002), Data management and Storage Group Leader at INFN-CNAF (the Italian Tier1 in LHC Computing Grid) from 2005 to 2011, and GARR Technical and Scientific Committee member from May 2009.

He is also INFN Tier1 manager since October 1st, 2011, and Director of CNAF since July 15, 2021.

Achille Felicetti achille.felicetti@pin.unifi.it

Achille Felicetti is an expert of knowledge representation and semantic web technologies and is active in the definition of ontologies and standards for Cultural Heritage. He is also responsible for the development and maintenance of digital services and infrastructures,



online collaborative platforms, databases, mapping services, and web-based OS applications.



Francesco Giacomini francesco.giacomini@cnaif.infn.it

First technologist at INFN-CNAF. His activity mainly concerns the design and implementation of software components both for distributed computing systems and for physics experiments.

Franco Niccolucci franco.niccolucci@pin.unifi.it

Franco Niccolucci is a former professor of the University of Florence and the Science and Technology in the archaeology Center of the Cyprus Institute. He has been the coordinator of the projects PARTHENOS, ARIADNE, CREATIVE CH, COINS, CHIRON, and 3D-ICONS. He is the author of about 100 publications in the domain of ICT applications to CH. He was the chair of CAA2004 and founder of the VAST conference series in which he co-chaired 2000, 2003, 2004, 2011, and 2012 editions.



Francesco Taccetti francesco.taccetti@fi.infn.it

First Technologist at INFN, expert in accelerators physics and applied physics on Cultural Heritages. He is national responsible for the INFN-CHNet, the INFN network devoted to the application of nuclear techniques to the Cultural Heritages.

He is Principal Investigator for the INFN in the European projects and MIUR-financed projects, such as 4CH, ORMA, Ariadne+, MACHINA, E-RIHS, and authored more than 80 papers published in international journals.

The technology transfer in Healthcare 4.0: the use of Machine Learning solutions for neurodegenerative diseases

Annamaria Demarinis Loiotile^{1,2}, Loredana Bellantuono², Francesco De Nicolò¹, Alfonso Monaco³, Sabina Tangaro^{3,4}, Nicola Amoroso^{3,5}, Roberto Bellotti^{2,3}

¹Dipartimento di Ingegneria Elettrica e dell'Informazione, Politecnico di Bari,

²Dipartimento Interateneo di Fisica, Università degli Studi di Bari Aldo Moro,

³Istituto Nazionale di Fisica Nucleare, Sez. di Bari, ⁴Dipartimento di Scienze del Suolo, della Pianta e degli Alimenti, Università degli Studi di Bari Aldo Moro,

⁵Dipartimento di Farmacia - Scienze del Farmaco, Università degli Studi di Bari Aldo Moro

Abstract. Artificial intelligence has the potential to revolutionize healthcare and machine learning approaches are able to be the catalyst for transformation of health systems to improve efficiency and effectiveness, warrant universal health coverage and improve outcomes. The paper shows some of the most relevant results emerged for the prediction of the most popular neurodegenerative diseases, and in particular a pipeline implemented for brain age prediction, exploiting deep learning. Some strategies to exploit these results are analyzed in order to create value for the society and the population, through a new service of Healthcare 4.0 based on the prediction of the most important neurodegenerative diseases, such as Alzheimer and Parkinson diseases.

Keywords. Technology transfer, Healthcare 4.0, machine learning, neurodegenerative diseases

Introduction

The health sector is particularly interesting for artificial intelligence (AI) applications, given the ongoing digitalization of health data and the promise for an improved quality of health and prevention (WHO Focus Group on Artificial Intelligence for Health).

By 2050, one in four people in Europe and North America will be over the age of 65 – this means the health systems will have to deal with more patients with complex needs. The management of these patients is expensive and requires systems to “shift from an episodic care-based philosophy to one that is much more proactive and focused on long-term care management” (Transforming healthcare with AI, 2020).

Artificial intelligence has the potential to revolutionize healthcare and help address some of these challenges, above all those related to neurodegenerative diseases which nowadays represent the challenge for Healthy and Active Ageing.

AI and machine learning will be the catalyst for transformation of health systems to improve efficiency and effectiveness, warrant universal health coverage (UHC) and impro-

ve outcomes (Panch, Szolovits et al., 2018).

The COVID 19 pandemic has highlighted the pressing need for improved data collection and exchange to better monitor and manage public health issues and health systems (Health at a Glance: Europe 2020).

With the creation of the European Health Data Space (EU 2020), the European Commission is currently developing a governance framework to promote a better use of health data, as well as a digital health infrastructure supporting such access. It will allow better use of data for health care, research, innovation and more evidence-based health policy-making (Health at a Glance: Europe 2020).

Fundamental transformation of health systems is critical to overcome these challenges and to achieve universal health coverage by 2030. Machine learning, the most tangible manifestation of artificial intelligence – and the newest growth area in digital technology – could be the catalyst for such a transformation (Panch, Szolovits et al., 2018).

The paper is organized as follow: Section 1 shows the important results emerged for the prediction of the most popular neurodegenerative diseases, and in particular a pipeline implemented for brain age prediction, exploiting deep learning; Section 2 presents some attempts for the valorization of these research results and their transfer in order to create impact on society. The conclusion of the paper is summarized in Section 3.

1. Deep learning approaches for neurodegenerative diseases

Neurodegenerative diseases affect a vast portion of the aging population and, of course, as life expectancy increases, this percentage is going to continuously grow. The onset of a neurodegenerative disease is usually related to several supportive features, among them an anomalous aging of the brain is particularly suitable for the early diagnosis pathological conditions. In brief, studying the structure, the organization and the functionality of the brain across the lifespan yields quantitative markers which can accurately discriminate the normal aging process from a pathological one. For example, measuring the difference between the chronological age of a patient and the brain age (usually referred to as “brain gap”) (Franke, Gaser, 2018; Bellantuono et al, 2021), estimated on the basis of specific age-related features such as cortical atrophy, allows the early detection of neurodegenerative processes.

The number of features which can be extracted to characterize a brain is huge, this is especially true when considering Magnetic Resonance Imaging (MRI) which capture the brain anatomy in millions of three dimensional voxels; therefore, the use of machine learning and deep learning approaches has become a necessity, both for computational and modelling purposes. Besides, several studies have demonstrated how such approaches can accurately predict the early onset of neurodegenerative diseases (Ashburner, 2007; Erus, Battapady et al, 2015; Konukoglu, Glocker et al, 2013; Sabuncu, Van Leemput, 2012). More recently, specific studies have explored the adoption of graph theory to model the brain laying the foundations for novel characterizations of the brain and the possibility to add graph-based features to the standard (anatomical-based) ones (Amoroso et al, 2018 a; Amoroso et al, 2018 b; Amoroso et al, 2018 c).

To manage the computational complexity of these strategies and fully exploit their informative content, the role of computer facilities has gained a paramount importance. A simple pipeline for the analysis of brain connectivity could require 24-48 hours of CPU time, therefore the investigation of a medical database which can include thousands of subjects would be just unfeasible using a laptop or even a workstation. On the contrary, the use of computer farms composed by thousands of CPU, the adoption (when possible) of parallel computing paradigms and, more recently, the technological development due to GPU processing, make these analyses affordable.

Finally, a not secondary aspect to mention is the possibility offered by learning algorithms to explain how the features used to feed the adopted models affect their accuracy. There is an increasing attention towards explainable machine learning, intended as an effort to make machine learning algorithms (in some cases considered and used as “black boxes” by physicians and clinical practitioners) more interpretable and easier to understand [Lombardi et al, 2020; Roscher et al, 2020; Binder et al, 2021]. This aspect is fundamental to deepen our understanding of pathological conditions and their etiology.

2. Transfer of research results

In order to generate an innovation process based on the concept of Value Based Healthcare and on the quadruple helix model, it is necessary that all of the following Entities work together: Regions; Large public and private hospitals, also organized in networks; Private Life Science companies; Academies, University and Research Centers. In particular, the “Academy” has a key role in developing and disseminating methods and tools to support implementation. The Universities can play as a catalyst and “super-partes” coordinator in the innovation process for Healthcare 4.0 (Harvard Business Review Italia, 2021).

Through Knowledge transfer or research results valorization the Universities promote the dissemination and use of new technologies developed at the research organizations in order to increase the impact of the research for all the stakeholders (Scanlan, 2018); in fact, the researchers could be able to design new solutions or discover new methodologies useful to mitigate long-lasting healthcare issues.

The pipeline implemented for neurodegenerative diseases using AI and deep learning approach, illustrated in the previous Section, could be valorized and transferred in order to become a tangible innovation for the society in terms of predicting neurodegenerative diseases. Thus, the research results could be exploited through a variety of complex channels including research conversion to IP and its patenting and licensing activity, collaborative research with private sector firms or contract research consulting with Public stakeholders, creation of academic start-ups or entrepreneurial entities, etc.

Similarly to what is done by the best universities in the world such as MIT, Stanford University, California Institute of Technology, Berkeley University, Oxford University, Harvard, Imperial College London etc, ... which use different ways and channels to enhance their research and create impact on the territory for the promotion of a positive and sustainable social impact, the authors, in collaboration with the Technology Transfer Office of their university, are planning the most fruitful strategy to exploit these results

in order to create value for the society and the population. In fact, these results could be transformed into a new service of Healthcare 4.0 based on the prediction of the most important neurodegenerative diseases, such as Alzheimer and Parkinson diseases.

3. Conclusion

In this paper some aspects related to Technology Transfer in Healthcare 4.0 are presented. In particular, according to the requested digitalization of health, the use of machine learning solutions for neurodegenerative disease prediction is analyzed in order to create benefits and impact on the society and the territory.

References

- Whitepaper for the ITU/WHO Focus Group on Artificial Intelligence for Health - https://www.itu.int/en/ITU-T/focusgroups/ai4h/Documents/FG-AI4H_Whitepaper.pdf
- EIT HEALTH (2020), Transforming healthcare with AI - The impact on the workforce and organisations - <https://thinktank.eithealth.eu/wp-content/uploads/2020/12/EIT-Health-and-McKinsey-%E2%80%93-Transforming-Healthcare-with-AI.pdf>
- Panch T., Szolovits P., Atun R. (2018), Artificial intelligence, machine learning and health systems, *J Glob Health* Vol. 8 No. 2, 020303
- OECD/European Union (2020), HEALTH AT A GLANCE: EUROPE 2020, https://www.oecd-ilibrary.org/social-issues-migration-health/health-at-a-glance-europe-2020_82129230-en
- Spazio europeo dei dati sanitari, https://ec.europa.eu/health/ehealth/dataspace_it
- Franke K., Gaser C. (2018), Ten years of BrainAGE as a neuroimaging bio-marker of brain aging: What insights have we gained? *Frontiers in neurology* 10, 789
- Bellantuono L. et al. (2021), Predicting brain age with complex networks: From adolescence to adulthood, *Neuroimage* 225: 117458
- Ashburner J. (2007), A fast diffeomorphic image registration algorithm, *Neuroimage* 38 (1), 95–113
- Erus G., Battapady H., Satterthwaite T.D., Hakonarson H., Gur R.E., Davatzikos C., Gur R.C. (2015), Imaging patterns of brain development and their relationship to cognition, *Cereb. Cortex* 25 (6), 1676–1684
- Konukoglu E., Glocker B., Zikic D., Criminisi A. (2013), Neighbourhood approximation using randomized forests, *Med. Image Anal.* 17 (7), 790–804
- Sabuncu M.R., Van Leemput K. (2012), The relevance voxel machine (RVoxM): a self-tuning Bayesian model for informative image-based prediction, *IEEE Trans. Med. Imaging* 31 (12), 2290–2306
- a. Amoroso N. et al. (2018), Complex networks reveal early MRI markers of Parkinson's disease, *Medical image analysis* 48: 12-24
- b. Amoroso N. et al. (2018), Deep learning reveals Alzheimer's disease onset in MCI subjects: results from an international challenge, *Journal of neuro-science methods* 302: 3-9

c. Amoroso N. et al. (2018), Multiplex networks for early diagnosis of Alzheimer's disease, *Frontiers in aging neuroscience* 10: 365

Lombardi A. et al. (2020), Association between Structural Connectivity and Generalized Cognitive Spectrum in Alzheimer's Disease, *Brain Sciences* 10.11: 879.

Roscher R. et al. (2020), Explainable machine learning for scientific insights and discoveries, *IEEE Access* 8: 42200-42216

Binder A. et al. (2021), Morphological and molecular breast cancer profiling through explainable machine learning, *Nature Machine Intelligence*: 1-12.

Harvard Business Review Italia (2021), White paper: Le sei priorità per implementare il Value Based Healthcare in Italia -- https://www.hbritalia.it/userUpload/Implementare_il_Value_Based_Healthcare_in_Italia.pdf

Scanlan J. (2018), A capability maturity framework for knowledge transfer, *Industry and Higher Education* Vol. 32(4) 235–244

Authors



Annamaria Demarinis Loiotile annamaria.demarinis@uniba.it

Graduated in Chemistry in 2006, she obtained a Master's degree in "Research Manager", an Advanced Training Course degree in "Research communication" and in "Innovation Broker". Since 2007, she has supported numerous national and international projects about environmental sustainability, health, innovation&technology transfer at the University of Bari. Today she works at the Innovation&Creativity Center of the University of Bari, as Research Manager and has started the PhD in Industry 4.0 at Polytechnic of Bari.

Loredana Bellantuono loredana.bellantuono@uniba.it

Dr. Loredana Bellantuono is an Assistant Professor in Applied Physics at the University of Bari. Her current research is focused on applications of complex network and machine learning methods to the analysis of complex systems in neuroscience, computational biology, social science, economy, and signal classification. She is a member of the Invitalia Operational Think Tank. After obtaining a PhD in Theoretical Physics, she worked as a visiting scientist at the Jagiellonian University in Krakow.



Francesco De Nicolò francesco.denicolo@poliba.it

Francesco De Nicolò received the Master Degree in Physics from the University of Bari discussing a thesis in Econophysics and Machine Learning. Currently he is a PhD student in "Industry 4.0" at the Polytechnic University of Bari – University of Bari. His research interests are in Machine Learning, Deep Learning, Complex Networks and their applications to Social Physics and Economy.

Alfonso Monaco Alfonso.Monaco@ba.infn.it

Graduated in Physics at the University of Bari and obtained also his PhD in Applied Physics, he is currently a technologist at INFN. His research interests concern fields of life science, astroparticle physics and scientific computing.



His research activity (over 200 publications, H-index=27) deals with the development of pattern classification techniques for data interpretation and signal approximation; the implementation of machine learning techniques for the study of complex systems in Grid/Cloud computing environment.



Sabina Tangaro sonia.tangaro@ba.infn.it

Associate professor of Applied Physics at the University of Bari, she focuses on the analysis of images and patterns gaining an expertise on image processing, computer vision and pattern recognition, machine learning, complex networks and coordinates some research projects founded by INFN and Italian Ministry of Research. Her work is focused on developing Computer Aided Detection software systems for Medical Applications on distributed databases. Her research activity includes explainable machine learning models.

Nicola Amoroso nicola.amoroso@uniba.it

Graduated in Physics at the University of Bari and PhD in Applied Physics, he has carried out research in the fields of Medical Physics, Econophysics and Earth Observation. He is currently a researcher at the Department of Pharmacy-Pharmaceutical Sciences, University of Bari. His interests are addressed to the modeling of complex data and the consequent design and implementation of quantitative/predictive analyses. He has co-authored 71 Scopus-indexed publications with a total h-index of 16 and 846 citations.



Roberto Bellotti roberto.bellotti@uniba.it

He has been working as a researcher and professor of Physics at the University of Bari since 1991 and he is currently full professor in applied physics and director of the Physics Department. He has been principal investigator of many research projects in medical physics and ICT. The areas of expertise extend to medical physics, signal and image processing, statistical pattern recognition, machine learning and complex network analysis. He is co-author of more than 300 peer-reviewed research, with > 20000 citations and h-index > 60.

Didattica Innovativa per l'Applicazione dell'Intelligenza Artificiale in contesti Industria 4.0

Daniele Mazzei, Riccardo Amadio, Daniele Atzeni

Dipartimento di Informatica, Università degli Studi di Pisa, Italia

Abstract. Il progetto PLANET4 ha l'obiettivo di colmare il gap di conoscenza tra la ricerca scientifica riguardante l'Intelligenza Artificiale (IA) e il Machine Learning (ML), e le sue applicazioni industriali come tecnologie abilitante per il paradigma dell'Industria 4.0. La ricerca accademica e l'insegnamento universitario sull'IA sono focalizzati principalmente sullo sviluppo e la creazione di modelli nuovi e complessi. Tuttavia, recenti ricerche hanno dimostrato la necessità di una decentralizzazione dell'IA, che consenta l'esecuzione degli algoritmi di analisi dei dati direttamente a bordo dei macchinari industriali (edge-processing). Il progetto PLANET4 mira a creare e sviluppare un corso di tipologia b-learning (misto), che possa promuovere il trasferimento di conoscenza tra mondo accademico e aziendale su questi temi con l'ambizione di dar vita quindi a una nuova categoria di esperti, caratterizzata sia da competenze tecniche sull'IA e il ML, sia da soft skills necessarie a gestire i cambiamenti introdotti dalla 4° rivoluzione industriale

Keywords. Intelligenza Artificiale, Cloud computing, Industria 4.0, Didattica

Introduzione

“Vogliamo potenziare l'Intelligenza Artificiale Europea mettendola al servizio delle persone e facendo sì che rispetti i diritti delle persone” – Ursula Von der Leyen, Presidente della Commissione Europea, durante la presentazione della strategia DigitalEU 19.02.2021. Il 14 Giugno 2018, è stata creata l'European AI Alliance (European Commission 2018) per assicurare la competitività dell'Unione Europea nel campo dell'Intelligenza Artificiale (IA). L'alleanza dichiarò la necessità di supportare il trasferimento di conoscenza tra mondo accademico e industria per facilitare l'utilizzo e la diffusione dell'IA e il Machine Learning (ML) nell'industria ma anche e soprattutto nelle piccole e medie imprese.

Vari studi (Microsoft Hypothesis Group 2020, Hassan 2018) prevedono un futuro dove miliardi di dispositivi di IoT (Internet of Things) saranno utilizzati all'interno delle fabbriche, consentendo la nascita e lo sviluppo di nuovi modelli di business incentrati sui dati e capaci di portare l'IA nell'industria (Kupper et al. 2018). Il paradigma del Digital Twin (gemello digitale) promosso dall'Industria 4.0 (I4.0) si immagina i dispositivi IoT come ponti in grado di collegare i dispositivi fisici e il Cloud, in cui l'IA risiede e dove vengono quindi prese le decisioni. Tuttavia, è innegabile che questo paradigma abbia delle limitazioni di scalabilità per via delle criticità legate ai consumi energetici (Talwalkar 2020) e alla necessità di una rete che fornisca una connessione veloce e al contempo sicura (Zhou et al. 2017).

Di conseguenza, gli esperti dell'I4.0 hanno appreso che è necessario muoversi verso u-

na decentralizzazione dell'architettura IoT, in cui l'IA è distribuita tra i dispositivi (paradigma comunemente chiamato Edge Computing, o semplicemente Edge) e il Cloud diventa il principale orchestratore e il gestore delle interfacce utente e dei cruscotti di monitoraggio (Shaw 2019). Al contrario, l'attuale tendenza all'interno del settore dell'IA, e soprattutto nelle comunità accademica focalizzata sul ML, è la promozione dello sviluppo e dell'insegnamento di architetture sofisticate, che richiedono un costo computazionale altissimo e di conseguenza non possono essere eseguite su dispositivi IoT (Bacciu et al. 2018).

L'unione di IA e IoT (chiamato AIoT) (Dialani 2019) offre un nuovo approccio alla computazione, che tenta di portare il processo di decisione tramite dati a livello locale. Il nuovo paradigma AIoT consente lo sviluppo di soluzioni I4.0 più sostenibili, scalabili, sicure e affidabili.

Le università hanno la quindi la possibilità e il dovere di crescere una nuova generazione di esperti nell'applicazione dell'AI on the Edge, partendo dalle necessità delle aziende e creando nuovi paradigmi riguardanti queste tematiche. Questo processo richiede di ideare, sviluppare e testare un nuovo metodo d'insegnamento per l'IA on the Edge per applicazioni industriali 4.0. Questo metodo dev'essere supportato da una serie di strumenti con i quali si possano apprendere le necessità delle aziende, formalizzare i problemi e identificare e implementare le migliori soluzioni attraverso tecniche di change-management, che agevolino la transizione verso queste nuove tecnologie.

L'obiettivo finale del progetto PLANET4 è quello di allenare una nuova generazione di tecnici, che siano contemporaneamente abili conoscitori di IA e ML e capaci di affrontare le applicazioni di queste tecnologie in ambito Edge. Il progetto cerca anche di rafforzare i legami tra le comunità di esperti di IA e ML, per accrescere e sostenere l'educazione degli studenti e le attività formative all'interno delle aziende tramite materiale formativo aggiornato. PLANET4 mira a costruire una struttura che faciliti l'uso delle tecnologie IA e ML nei settori produttivi. I futuri ingegneri e tecnici hanno bisogno di un nuovo "super-paradigma" di progettazione, capace di gestire sia le opportunità sia la complessità e le ambiguità dell'IA nell'I4.0.

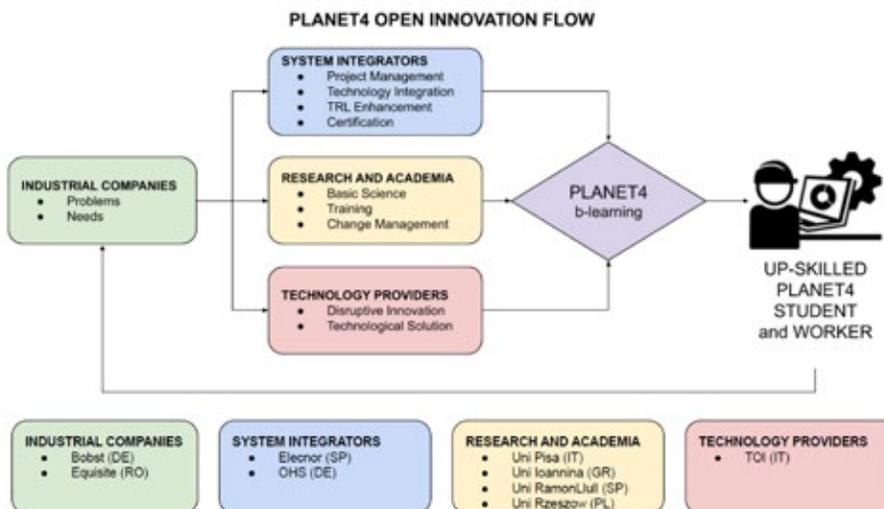


Fig. 1
Schema organizzativo e flusso innovativo di PLANET4

1. Metodologia

Lo scopo di PLANET4 è quello di creare e realizzare un corso b-learning extra-curriculare innovativo sul tema dell'AI on the Edge per applicazioni Industria 4.0, riservato a studenti universitari e dipendenti di aziende manifatturiere. La quarta rivoluzione industriale ha portato le aziende ad affrontare bisogni e problematiche difficili da risolvere utilizzando esclusivamente le tecnologie a loro disposizione o metodi ben conosciuti. Per questo motivo, è necessario incentivare il trasferimento di conoscenza tra settore industriale, ricerca scientifica e il mondo dell'impresa. PLANET4 mira ad ampliare questo processo, includendo soluzioni basate su AIoT.

Come mostrato nella Figura 1, PLANET4 mette in comunicazione azienda, fornitori di tecnologie, integratori di sistemi ed il mondo accademico e della ricerca per costruire un processo di innovazione più sostenibile, scalabile e resiliente.

Le attività di PLANET4 possono essere divise in tre fasi, come mostrato in Figura 2:

- Fase 1: questa fase si concentra sull'analisi e sul raccoglimento di dati relativi sia alle necessità e ai problemi delle aziende, sia alle metodologie e alle ricerche riguardanti tecnologie e soluzioni utilizzate per aggredire problemi tipici dell'IA applicata all'I4.0. I risultati di questa analisi verranno organizzati in una tassonomia, il cui scopo è quello di identificare le relazioni esistenti e mancanti tra necessità aziendali e soluzioni tecnologiche. Inoltre, in questa fase verranno individuate le più appropriate metodologie di Innovation e Change Management per l'I4.0.
- Fase 2: lo scopo di questa fase è l'analisi delle metodologie di insegnamento più utilizzate riguardo alle competenze tecniche e le soft skills necessarie per l'applicazione di modelli di IA nell'I4.0. I risultati di questa analisi consentiranno la preparazione e la pianificazione del corso.
- Fase 3: in questa fase, le metodologie di insegnamento vengono testate e validate. Il corso sarà suddiviso in una parte teorica, sviluppata in e-learning e riguardante l'utilizzo dell'AI on the Edge e delle soft skills, e due workshop: 1) Metodologie di Innovation e Change Management per l'I4.0 e 2) Apprendimento, Classificazione e Risoluzione delle sfide relative dell'I4.0.

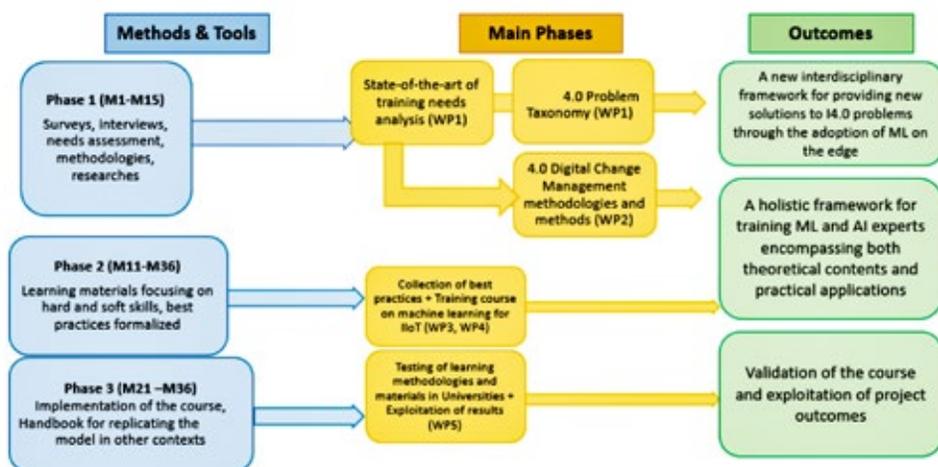


Fig. 2
Fasi principali
di PLANET4

Al termine del progetto, alcuni dei casi analizzati e risolti durante il corso (Fase 3, workshop 2) saranno pubblicati nel sito web del progetto per fornire ad altre organizzazioni esempi e sfide basati su problemi reali dell'ambito industriale.

2. Conclusioni

In questo articolo è stato introdotto il progetto PLANET4, il cui scopo è la pianificazione e l'implementazione di un corso extra curriculare rivolto a studenti e lavoratori che consenta la creazione di nuove figure professionali che possano giocare un ruolo chiave durante la transizione al paradigma dell'Industria 4.0. Questo processo di trasferimento tecnologico è oggi più che mai fondamentale dal momento che con il PNRR (Piano Nazionale di Ripresa e Resilienza) l'Italia si trova davanti ad una delle più grandi sfide di innovazione del comparto produttivo che si sia mai affrontata dal secondo dopoguerra.

Tramite un approccio innovativo, questo progetto permette inoltre il raccoglimento e la formalizzazione delle necessità e dei problemi affrontati dalle aziende e la loro classificazione in una tassonomia che può fungere da strumento facilitante del processo di trasferimento tecnologico e di conoscenze relativo all'applicazione dell'intelligenza artificiale in applicazioni industria 4.0

3. Ringraziamenti

Questo lavoro è stato in parte finanziato dal programma Erasmus+, Knowledge Alliances, Application No 621639-EPP-1-2020-IT-EPPKA2-KA, PLANET4 (Practical Learning of Artificial Intelligence on the Edge for industry 4.0)

Riferimenti bibliografici

D. Bacciu, S. Chessa, C. Gallicchio, A. Micheli (2017), On the need of machine learning as a service for the Internet of Things, in Proceedings of the 1st International Conference on Internet of Things and Machine Learning, pp 1-8.

P. Dialani (2019), IoT and AI at the Edge Creating Artificial Intelligence of Things, available at <https://www.analyticsinsight.net/iot-and-ai-at-the-edge>.

European Commission (2018), Artificial Intelligence for Europe, available at <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?url=CELEX:52018DC0237&from=EN>.

Q. F. Hassan (2018), The Industrial Internet of Things

D. Kupper, M. Lorenz, K. Kuhlmann, O. Bouffault, J. Van Wyck, S. Kocher, J. Schlager (2018), AI in the Factory of the Future, available at <https://www.bcg.com/it-it/publications/2018/artificial-intelligence-factory-future>.

Microsoft Hypothesis Group (2020), IoT Signals Edition 2, available at <https://azure.microsoft.com/en-gb/iot/signals>.

G. Shaw (2019), The Future Computed AI and Manufacturing, available at https://accenterp.com/wp-content/uploads/2019/05/Microsoft_TheFutureComputed_AI_MFG_Final_.pdf%EF%BF%BD.

A. Talwalkar (2020), AI in the 2020s Must Get Greener - and Here's How, available at

<https://spectrum.ieee.org/energy-efficient-green-ai-strategies#toggle-gdpr>.

J. Zhou, Z.Cao, X.Dong, A.V. Vasilakos (2017), Security and Privacy for Cloud-based IoT: Challenges, IEEE Communications Magazine, vol 55, no. 1, pp 26-33.

Autori



Daniele Mazzei daniele.mazzei@unipi.it

Laureato in Ingegneria Biomedica, Dottorato in Robotica, Automazione e Bioingegneria presso l'Università di Pisa. Ricercatore presso il Dipartimento di Informatica dell'Università di Pisa, Cofondatore e Chief Innovation Officer di Zerynth. La ricerca di Mazzei verte sullo studio e applicazione di tecniche di intelligenza artificiale in contesti industriali e sullo sviluppo di architetture IOT a supporto del paradigma 4.0

Riccardo Amadio r.amadio@studenti.unipi.it

Laureando in Computer Science e Artificial Intelligence all'Università di Pisa.. Ricercatore borsista al Dipartimento di Informatica dell'Università di Pisa. La ricerca di Amadio veerte sullo studio e applicazione di Intelligenza Artificiale applicate all'ambito industriale e IoT.



Daniele Atzeni daniele.atze@gmail.com

Laureato in Matematica presso l'Università di Genova ed in Data Science and Business Informatics presso l'Università di Pisa, ricercatore all'interno del Dipartimento di Informatica dell'Università di Pisa. La ricerca di Atzeni si occupa dello studio degli algoritmi di Intelligenza Artificiale e Machine Learning applicati in contesti reali, soprattutto in ambito industriale ed IoT.

L'intelligenza artificiale contro la violenza di genere

Claudia Villante, Gabriele Paone, Maria Giuseppina Muratore

Istat

Abstract. Obiettivo dell'attività di diffusione dell'Istituto Nazionale di Statistica è garantire l'accesso ai risultati delle rilevazioni statistiche utilizzando tecnologie aggiornate. Allo stato attuale infatti l'ISTAT ha dedicato un sito web al tema della violenza di genere (<https://www.istat.it/it/violenza-sulle-donne>), dove sono riportati dati e informazioni sul fenomeno che sono in continuo aggiornamento. Il sistema informativo "violenza sulle donne" è stato realizzato in collaborazione con il Dipartimento per le Pari Opportunità della Presidenza del Consiglio dei Ministri e risponde all'esigenza di monitorare il fenomeno nei suoi vari aspetti, in linea con le indicazioni della Convenzione di Istanbul. Il sistema produce inoltre dati e indicatori necessari a verificare il raggiungimento degli obiettivi previsti dall'Agenda 2030 – Goal 5 "Raggiungere l'uguaglianza di genere ed emancipare tutte le donne e le ragazze"

Keywords. Gender based Violence, gender stereotype, big data, sentiment analysis, emotion analysis

Introduzione

L'obiettivo del progetto è quello di arricchire il patrimonio informativo statistico oggi disponibile con indagini sperimentali come la sentiment e l'emotion analysis avvalendosi di tecnologie all'avanguardia come il machine learning.

Allo stesso tempo lo scopo è quello di rafforzare l'efficacia e la qualità dei dati prodotti, predisponendo analisi, raccolte e bonifica di dati agli organi di governo e a tutti i soggetti pubblici e privati coinvolti nelle azioni di contrasto alla violenza di genere.

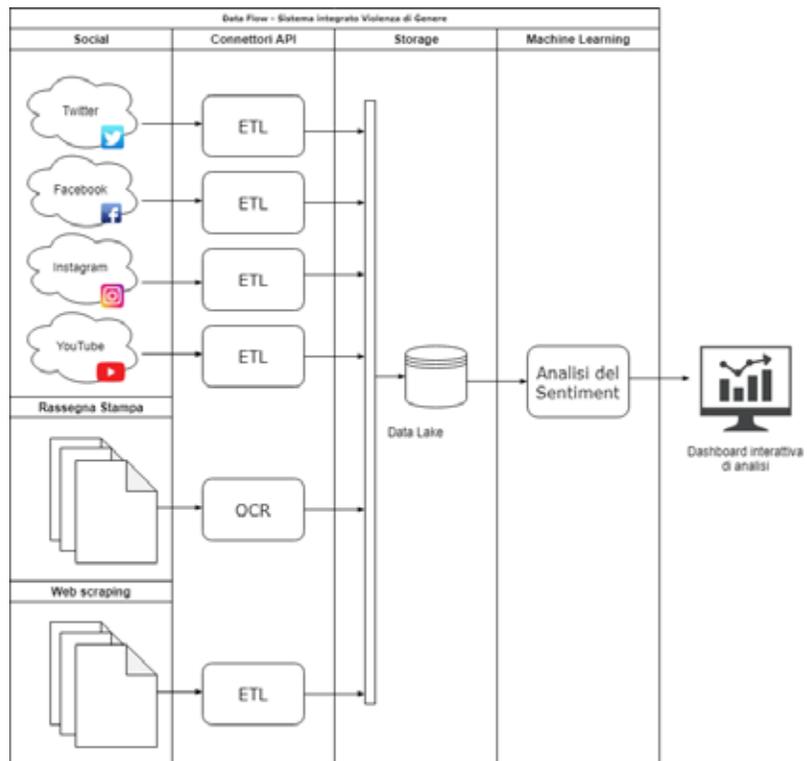
1. Approccio architetturale

Il primo passo è dunque quello di integrare le varie fonti. Il servizio si avvale di una soluzione che, a partire dai requisiti espressi dagli statistici dell'Istituto, abilita la raccolta delle informazioni sui diversi canali individuati e la definizione di viste-ambiti tematici (insights) di interesse, la selezione ed elaborazione di KPI specifici, l'addestramento degli algoritmi di calcolo e la produzione della reportistica per la fruizione dei risultati da parte degli utenti per una completa e certificata Web and Social Analysis.

È stata realizzata una dashboard con quattro quadri di analisi (insight) e una psicografica, basata su prodotti open source, per la visualizzazione e la navigazione delle informazioni. Il flusso completo alla base dell'erogazione del servizio viene rappresentato sinteticamente nella figura seguente.

In risposta alla sicurezza ed alla privacy dei dati (Regolamento UE – GDPR), i principali Social Network (Twitter, Facebook, Instagram, YouTube) stanno revisionando le proprie policy. La concreta compatibilità d'uso andrà quindi valutata in funzione delle specifiche

Fig. 1
Flusso dati



esigenze dell'Istituto ed in relazione alle policy una volta consolidate.

La configurazione dei canali considera un'attività di analisi preventiva finalizzata a:

- selezionare i canali più idonei (ovvero dove andare a raccogliere i dati)
- definire gli ambiti, temi, soggetti (ovvero cosa raccogliere)

L'accuratezza della configurazione dei canali, sotto il profilo qualitativo, risulta essenziale in quanto permette raccolte più mirate, evita acquisizioni di moli di dati scarsamente, o per nulla, significativi per le finalità dell'Istituto, e abilita l'ottenimento di risultati di maggior qualità nell'erogazione del servizio.

2. Gli Insights

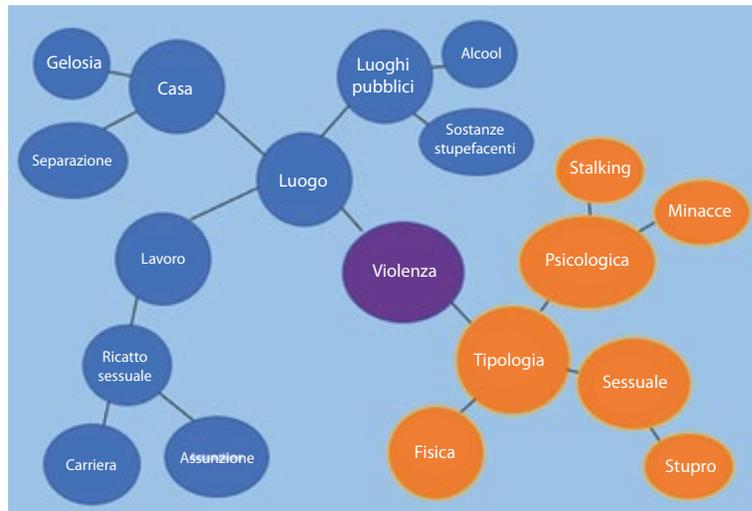
I contenuti acquisiti vengono analizzati, classificati ed organizzati in macro-ambiti. A puro titolo di esempio, la figura seguente riporta alcune rappresentazioni possibili.

3. La Dashboard interattiva

I principi architetturali precedentemente riassunti sono stati declinati nella realizzazione di una Dashboard interattiva che permette la configurazione e l'utilizzo di KPI standard e personalizzabili secondo i quali è possibile customizzare dashboard analitiche (es. analisi dei trend, andamento KPI di sintesi, analisi integrata dei contenuti).

A titolo di esempio viene di seguito fornito un elenco dei principali KPI implementabili per la fonte di tipo Social: - termini ricorrenti - hashtag frequenti - utenti più attivi - utenti più menzionati - distribuzione linguistica - distribuzione like (Instagram) - distribuzione

Fig. 2
Gli Insights



reply - distribuzione commenti – sentiment - trend temporali e picchi. Interessante il fatto che sia in fase di raccolta sia in fase di visualizzazione, sarà possibile analizzare i temi utilizzando dizionari personalizzati dall'utente.

4. La metodologia per la misurazione del Sentiment e dell'Emotion

Il servizio di sentiment analysis utilizza metodologie che si fondano su architetture di Deep Learning (Artificial Intelligence). La precisione dei risultati in termini di sentiment e di emotion è assicurata, oltre che dalla qualità intrinseca della metodologia, proprio dall'attività condotta in fase di addestramento, utilizzando tecniche di Active learning.

A partire da un ambito-tema o dal soggetto di interesse per l'Istituto, l'addestramento del motore (Machine Learning) si basa sulla disponibilità di un campione di testi certificati in termini di risposta sentiment (valutazioni positive, negative, neutre) ed emotion (gioia, sorpresa, amore, tristezza, rabbia, neutro); l'accuratezza delle classificazioni viene arricchita e perfezionata tramite l'impiego di opportuni cataloghi creati ad-hoc, secondo le specifiche esigenze del contesto di interesse, in questo caso la violenza di genere e gli stereotipi di genere.

L'addestramento è propedeutico per la successiva definizione del calcolo del sentiment e dell'emotion e, dunque, dell'erogazione del servizio. L'analisi del sentiment e dell'emotion risulta tanto più accurata quanto più ampio è il campione utilizzato in fase di addestramento (casistiche di classificazione). Al fine di perfezionare e rendere più affidabile i risultati delle analisi si prevedono fasi successive di tuning.

Conclusioni

L'analisi è iniziata a marzo 2020 e da questa data la piattaforma è costantemente aggiornata in tempo reale. Sulla base della proposta metodologica si prevede la produzione di una statistica sperimentale con cadenza trimestrale volta a restituire informazioni e colmare un gap informativo importante, offrendo informazioni inedite sul clima/mood che si crea intorno ai principali eventi riconducibili al tema della violenza contro le donne e de-

gli stereotipi di genere. L'obiettivo quindi è quello di leggere attraverso indici di positività o negatività come vengono veicolati i messaggi social sul fenomeno oggetto di analisi, ma anche di osservare come lo stesso fenomeno della violenza di genere, al pari di molti altri fenomeni sociali (come ad esempio il cyberbullismo) si espande nelle sue diverse forme digitali. Nuove estrinsecazioni del fenomeno richiedono, infatti, l'elaborazione di nuovi strumenti di analisi che siano in grado di coglierlo. Con una uscita periodica di reports inoltre si intende perseguire un ulteriore obiettivo di tipo culturale che attiene alla volontà di tenere alta l'attenzione dell'opinione pubblica sugli stereotipi di genere e l'immagine sociale della violenza, in linea con le raccomandazioni internazionali e con il raggiungimento degli obiettivi dell'Agenda 2030.

Riferimenti bibliografici

Bergvall, V. (1999). Toward a comprehensive theory of language and gender. *Language in Society*, 28(2), 273– 293. Retrieved from www.jstor.org/stable/4168929.

Berzal, Fernando & Matín, Nicolfás. (2002). Data mining: concepts and techniques by Jiawei Han and Micheline Kamber. *ACM SIGMOD Record*. 31. 66-68. 10.1145/565117.565130.

Bing Liu and Ian Lane. 2016. Attention-based recurrent neural network models for joint intent detection and slot filling. In *Interspeech 2016*, pages 685–689.

Bolasco S., 1999, "Analisi multidimensionale dei dati", Carocci Editore.

Bonadiman D., Castellucci G., Favalli A., Romagnoli R. and Moschitti A., (2017) *Neural Sentiment Analysis for a Real-World Application*, DOI: 10.4000/books.aaccademia.2357, Accademia University Press, Torino

Chakraborty, Goutam. (2014). *Analysis of Unstructured Data: Applications of Text Analytics and Sentiment Mining*.

Charles J. Fillmore. 1985. Frames and the semantics of understanding. *Quaderni di Semantica*, 6(2):222– 254.

Autori

Claudia Villante claudia.villante@istat.it

Dottore di ricerca in Sistemi sociali e analisi delle politiche pubbliche, dal 2018 e' primo ricercatore ISTAT, presso la Direzione Centrale Statistiche Sociali e Censimento della popolazione. E' stata policy officer presso ANPAL e ricercatore presso INAPP (ex-Isfol) occupandosi del monitoraggio e valutazione delle politiche pubbliche, in chiave di genere. Attualmente e' esperta nazionale distaccata presso EIGE (European Institute for Gender Equality).

Gabriele Paone gpaone@istat.it

Nasce nel 1975, è un ingegnere informatico e attualmente lavora come IT Senior Architect presso la Direzione informatica dell'ISTAT, inoltre è responsabile tecnico degli applicativi d'area "Salute, assistenza, previdenza, giustizia, lavoro, istruzione e formazione" all'interno dell'iniziativa "Sistemi Informativi per le statistiche sociali e il censimento della popolazione".

Fa parte della Commissione informatica dell'Ordine degli Ingegneri di Roma.

Maria Giuseppina Muratore muratore@istat.it

dottore di ricerca in Metodologia della Ricerca Sociale, primo ricercatore Istat dal 2004, è responsabile del Sistema Informativo sulla violenza di genere, nato dall'Accordo tra l'Istat e il Dipartimento per le Pari Opportunità presso la Presidenza del Consiglio nel 2017, delle rilevazioni sui Centri antiviolenza e le Case rifugio, dell'indagine sulla sicurezza delle donne, dell'indagine sulla sicurezza dei cittadini. Esperta su crime, crime classification e gender based violence è stata parte di molte Task Force e gruppi di lavoro presso le Nazioni Unite (UNECE, UNODC, UNWOMEN, UNSD), il Centre of Excellence on Crime, la Commissione Europea, EUROSTAT, il FRA (European Fundamental Rights Agency) e EIGE.

Infrastrutture digitali per i dati della ricerca: adozione e personalizzazione del Research Data Management System all'Università Milano-Bicocca

Bonaria Biancu, Alessandro Andretto, Paolo Brambilla

Università degli Studi di Milano-Bicocca

Abstract. BOARD (Bicocca Open Archive Research Data) è il Research Data Management System che l'Università di Milano-Bicocca mette a disposizione di docenti e ricercatori per il deposito e la pubblicazione dei dati della ricerca. Frutto di un'attenta analisi comparativa tra i prodotti esistenti sul mercato, da luglio 2020 la piattaforma è soggetta a intense implementazioni e personalizzazioni, non solo per soddisfare le esigenze dell'istituzione ma anche per contribuire a restituire un servizio valido all'intera comunità di prodotto

Keywords. Open Science, Research Data, Research Data Management, Principi FAIR

Introduzione

La policy sull'Open Science approvata nel novembre 2019 dall'Università Milano-Bicocca poggia su tre cardini: Open Access per le pubblicazioni scientifiche, Open data per i dati della ricerca, Open Infrastructure per l'utilizzo in modalità open delle infrastrutture di ricerca dell'Ateneo.

Per la gestione dei dati della ricerca, tra le prime azioni che l'Ateneo ha posto in essere per dare seguito a quanto stabilito nella policy, vi sono state l'assunzione di una figura dedicata di data steward e l'adozione di una piattaforma per la gestione, pubblicazione e disseminazione dei dati.

Per la scelta del Research Data Management System (RDMS) si è cercata una soluzione che contemperasse le esigenze concrete dei ricercatori dell'Ateneo, le indicazioni della policy sulla Scienza Aperta e la gestione tecnico-organizzativa del processo di pubblicazione degli open data. L'adozione del RDMS è stata preceduta da una lunga e articolata fase di analisi, nella quale sono state messe a confronto le principali piattaforme di Research Data Management attualmente presenti sul mercato. Assodato che tutte le piattaforme erano compliant ai principi FAIR, in fase di analisi è stato dato molto peso alla presenza di determinate caratteristiche tecniche, valutate come imprescindibili per la scelta finale.

Il risultato di questa approfondita indagine ha portato all'impiego e allo sviluppo di BOARD (Bicocca Open Archive - Research Data; Url: <https://board.unimib.it/>), il RDMS fornito a Bicocca in modalità Software as a Service da Elsevier.

1. BOARD: caratteristiche della piattaforma

La scelta del RDMS è stata determinata tenendo bene a mente alcune caratteristiche fon-

damentali, considerate dei must have per la piattaforma dei dati di Bicocca.

È innanzitutto importante ribadire che la piattaforma è FAIR compliant: al dataset pubblicato vengono associati un DOI e un numero di versione, in modo da renderlo immediatamente citabile e da poter tenere traccia di eventuali, successive modifiche. Le 15 licenze a disposizione (preset sulle Creative Commons ma possibilità di selezionarne altre, sempre in ambito Open Access/Open Source) permettono inoltre di qualificare la disseminazione dei dati in modo corretto.

La pubblicazione dei dataset è soggetta a un processo di validazione e approvazione, necessario perché i data steward possano verificare la corretta implementazione dei principi FAIR e della policy di Ateneo sull'Open Science, nonché l'allineamento con quanto dichiarato in un eventuale Data Management Plan (DMP).

Altri punti qualificanti per la scelta del repository sono stati l'indicizzazione e la ri-pubblicazione dei dataset di Bicocca già presenti sulle principali piattaforme di disseminazione dei dati della ricerca: attività, queste, che rispondono all'esigenza di non interferire con le abitudini dei ricercatori, ormai invalse nei vari ambiti disciplinari. È necessario evitare di depositare più volte gli stessi dataset, sia per l'ovvio motivo di non moltiplicare i punti di accesso ai dati creando nel tempo inconsistenze, sia per non dover chiedere lo sforzo di ri-caricare dati già depositati su altre piattaforme. In BOARD sono così ricercabili i dataset già resi disponibili altrove (circa 1700 a settembre 2021), mediante la pubblicazione dei loro riferimenti e dei metadati principali, e un puntatore alle sorgenti di provenienza. Le risorse sono state inoltre organizzate in collezioni dipartimentali.

Per quanto riguarda il caricamento dei dati sul repository istituzionale, si è agito sulla personalizzazione del workflow di deposito, a partire dalle impostazioni relative ai connettori verso i servizi di storage forniti dai principali cloud provider. Essenziale è stato l'interfacciamento con i connettori di Google Drive e Microsoft OneDrive, dal momento che Bicocca ha sposato a livello istituzionale l'adozione sia della Google Suite sia di Office365. Ugualmente importante l'integrazione con Microsoft Azure, che risponde all'esigenza di offrire, a quei ricercatori che già utilizzano risorse e laboratori virtuali in cloud, una modalità semplice e immediata per trasferire i propri dati dall'infrastruttura virtuale di Azure a BOARD. Sono inoltre garantiti i connettori a DropBox e Box, ed è supportato il protocollo WebDav per il caricamento di file da server remoti.

L'accesso a BOARD è inoltre integrato con il sistema di autenticazione federata di Ateneo. Oltre ai comuni vantaggi di un accesso federato, questa integrazione permette di agganciare in automatico i profili degli utenti affiliati a Bicocca ed assegnare loro i privilegi del repository istituzionale (100 GB di spazio per dataset, metadati customizzati, affiliazione automatica all'istituzione e al dipartimento).

Infine, ai tempi in cui è stata svolta l'analisi comparativa, il servizio di Elsevier era l'unico, tra i vari competitor analizzati, che includesse uno spazio di archiviazione per la conservazione perpetua di dati e metadati. La long term preservation è garantita tramite un accordo con il DANS (Data Archiving and Networked Services) olandese e permette a Bicocca di evitare la completa dipendenza dal fornitore.

2. Integrazioni e sviluppi

L'adozione di un RDMS richiede un processo di miglioramento continuo, al fine di aumentare la qualità dei servizi già disponibili e di implementare nuove funzionalità con l'obiettivo di rendere il sistema funzionale ed interoperabile, mettendo al centro le esigenze dei ricercatori.

Proprio ai fini dell'interoperabilità è iniziato un percorso che avrà come obiettivo finale l'integrazione tramite web service REST con le fonti dati di Ateneo, in primis le anagrafiche utenti di IRIS (il CRIS di Bicocca). Un primo risultato evidente in questa direzione è la funzionalità di affiliazione automatica, che associa di default il contributor del dataset all'istituzione e al dipartimento di appartenenza. Anche in questo caso è valso il principio di non interferire con le abitudini di pubblicazione dei ricercatori, abituati in IRIS a trovare già precompilata la propria affiliazione a livello dipartimentale. Sempre sull'interoperabilità, sono a disposizione le API (e la relativa documentazione) per interrogare i servizi e ottenere informazioni previa autenticazione.

Altro aspetto non secondario è rappresentato dai servizi complementari: è in sviluppo un sistema di reportistica che fornisce pubblicamente l'analisi dei dati di utilizzo di BOARD, mentre delle specifiche aree riservate permettono a utenti e amministratori di approfondire il grado di utilizzo e di disseminazione delle risorse depositate.

Il mantenimento di un alto standard di qualità dei metadati rimane un punto fondamentale: mentre per gli schemi da adottare si è cercato di attingere, per quanto possibile, allo standard Dublin Core, vi è una costante ricerca volta al miglioramento delle informazioni fornite. Tra i metadati personalizzati per Bicocca sono presenti le classificazioni ERC e SSD, mentre si sta lavorando per una possibile integrazione di tassonomie per la classificazione formale dei ruoli dei contributor al dataset.

Ultimo, ma non meno importante, è un generale approccio verso l'internazionalizzazione: BOARD è già registrato in registri europei di repository dei dati, come FAIRsharing (<https://doi.org/10.25504/FAIRsharing.DDotIl>) e Re3Data (<http://doi.org/10.17616/R31NJMUX>). In ottica EOSC, invece, l'obiettivo è di diventare data provider dell'infrastruttura europea.

3. Conclusioni: sostenibile e digitale

Guardando alle implementazioni realizzate fino ad oggi e agli sviluppi futuri, è chiaro come il soddisfacimento delle necessità degli utenti rimanga il punto centrale per lo sviluppo di BOARD. Questo approccio si è tradotto in soluzioni pratiche, frutto anche di un confronto iniziato a settembre 2020 con i ricercatori di Bicocca mediante la partecipazione a focus group creati per migliorare la user experience della piattaforma.

Proprio in ragione di questa attenzione costante verso i docenti e i ricercatori che dovranno effettivamente gestire e pubblicare i dati, si è prestata attenzione a soddisfare le esigenze di enti finanziatori e riviste, ormai orientati sempre più a richiedere di default il deposito dei dati della ricerca in repository qualificati.

La predilezione per l'opzione SaaS incontra il favore delle direttive AgID per le quali la soluzione di Elsevier è qualificata. Proprio questa qualificazione - voluta fortemente da Bi-

cocca - assicura che il servizio di Elsevier sia sviluppato e fornito secondo criteri di affidabilità e sicurezza, considerati necessari per i servizi digitali pubblici. In aggiunta, BOARD garantisce un ulteriore livello di sicurezza sulla disponibilità delle risorse grazie alle API, che permettono di esportare massivamente dati e metadati in qualsiasi momento.

Vale infine la pena di evidenziare un ritorno sugli investimenti che l'adozione di questa piattaforma ha fin qui generato: dal punto di vista dei costi diretti, gli sforzi implementativi del gruppo di lavoro di Bicocca, in collaborazione col team di Elsevier, hanno reso la piattaforma accogliente, aperta, strutturata ma flessibile ed interoperabile, e predisposta alla long term preservation. Di questo gioverà l'intera comunità di prodotto, che oggi trova una soluzione sicuramente più funzionale di quella presentata nel 2020. Dal punto di vista dei costi indiretti, non meno importanti, gli sforzi portati avanti mirano ad evitare inutili duplicazioni dei dati, in modo da ridurre quel caos informativo in cui è purtroppo molto facile incappare.

Autori



Bonaria Biancu bonaria.biancu@unimib.it

Bonaria Biancu lavora in università dal 2002. Si occupa di gestione dei sistemi per la ricerca e dei servizi web, e di digitalizzazione dei processi di business. Dal 2020 è componente del Consiglio di Amministrazione dell'Università Milano-Bicocca.

Alessandro Andretto alessandro.andretto@unimib.it

Alessandro Andretto è il responsabile dell'ufficio Sistemi Integrati per la Ricerca presso i Sistemi Informativi di Bicocca e si occupa della gestione del CRIS di Ateneo, Open Science, campagne di valutazione e business intelligence per la ricerca.



Paolo Brambilla paolo.brambilla2@unimib.it

Paolo Brambilla è Data Specialist presso l'Università di Milano-Bicocca. Si occupa principalmente di Open Science ed è amministratore della piattaforma BOARD. In precedenza ha lavorato presso diversi enti e istituzioni per lo sviluppo di politiche sui dati aperti.

Il percorso di adozione del paradigma Open Science all'INGV

Mario Locati, Francesco Mariano Mele, Placido Montalto, Vincenzo Romano, Valentino Lauciani, Roberto Vallone, Stefano Cacciaguerra

Istituto Nazionale di Geofisica e Vulcanologia (INGV)

Abstract. L'Istituto Nazionale di Geofisica e Vulcanologia (INGV) svolge attività di ricerca scientifica nell'ambito delle Scienze della Terra e attività all'interno del sistema nazionale di Protezione Civile grazie all'uso delle sue numerose reti di monitoraggio multiparametrico dislocate sull'intero territorio nazionale. L'INGV intende adottare il paradigma dell'Open Science, ovvero un approccio alla Ricerca basato su una più veloce ed efficiente condivisione dei dati per favorire nuovi scenari multidisciplinari inesplorati. In questo contesto, l'INGV si è dotato di una Politica dei Dati che predispone gli strumenti gestionali necessari all'attuazione dell'Open Science e dei principi FAIR e demanda il coordinamento di queste attività all'Ufficio Gestione Dati. Ogni azione è intrapresa considerando i processi scientifici, tecnologici e gestionali che gravitano intorno al dato, per crearlo, mantenerlo, condividerlo, conservarlo, tracciarne la provenienza e l'uso, e infine misurarne l'impatto

Keywords. Open Science, Data Policy, Open Data, FAIR Data, EOSC

Introduzione

Lo scambio di dati nell'ambito delle Scienze della Terra è parte integrante di questa disciplina dato che i fenomeni naturali ignorano i confini amministrativi. Da sempre le comunità attive nel settore condividono informazioni beneficiando dell'evoluzione tecnologica perché i dati raccolti, analizzati e condivisi crescono in quantità e sono sempre più complessi. Già negli anni '80 del secolo scorso INGV, allora chiamato ING, realizzò la prima rete di sismica attiva con qualche decina di stazioni su scala mediterranea. Oggi INGV gestisce reti di monitoraggio multi-parametrico con diverse centinaia di stazioni dislocate sull'intero territorio nazionale e si avvale inoltre di dati da altre reti di monitoraggio gestite da altre organizzazioni come il Dipartimento di Protezione Civile o reti satellitari gestite dall'Agenzia Spaziale Europea e dalla NASA.

La standardizzazione dei processi per condividere i dati dell'INGV è una sfida complessa che copre varie aree disciplinari condotte in una struttura distribuita geograficamente su sette sezioni principali (Milano, Bologna, Pisa, Roma, Napoli, Palermo e Catania) e ventuno altre sedi distaccate sparse in tutta Italia.

È importante sottolineare come le informazioni raccolte sui fenomeni naturali osservati siano estremamente preziose perché per loro stessa natura uniche e irripetibili. È quindi fondamentale adottare soluzioni di conservazione dei dati con un orizzonte temporale di lungo periodo.

La ricerca condotta all'INGV adotta il paradigma dell'Open Science (Nielsen 2011) dal

momento che uno dei cardini della propria attività è la condivisione il più possibile libera, semplice, veloce ed efficiente della conoscenza, sfruttando le moderne tecnologie di comunicazione. La condivisione delle informazioni deve permettere a chiunque un semplice utilizzo dei dati nel maggior numero possibile di ambiti scientifici a potenziale beneficio di tutta la comunità.

La fase preliminare

Per attuare i principi dell'Open Science servono soluzioni concrete che possano essere adottate dai ricercatori minimizzando eventuali impatti negativi sulle loro attività quotidiane (Meadows 2018). Servono pratiche efficaci che permettano di migliorare il lavoro dei ricercatori, ad esempio per meglio mettere a frutto l'enorme e complesso patrimonio di conoscenza oggi disponibile ma spesso di difficile accesso e uso. A prescindere dagli strumenti adottati, il loro successo passa da un inquadramento amministrativo-gestionale che ne stabilisca finalità, tempistiche, costi, ruoli e responsabilità la cui evoluzione potrà comportare il cambiamento di processi gestionali consolidati.

L'adozione dei principi dell'Open Science è ovunque in fase sperimentale e per la sua attuazione è difficile trovare sia esempi pratici da esaminare, sia l'ingaggio di personale qualificato. Al fine di avviare una sperimentazione, INGV si è avvalso di figure professionali già presenti nell'istituto che hanno seguito un percorso auto-formativo recuperando informazioni e documentazione dalle fonti più disparate di ambito scientifico, normativo e informatico.

Fin dal 2012-2013 all'INGV si iniziò a discutere di quali fossero le possibili strade da percorrere, discussioni condotte sia all'interno dell'istituto, sia con altre organizzazioni europee nell'ambito di progetti come FP7 EUDAT e FP7 European Plate Observing System (EPOS). Queste discussioni culminarono nel 2013 con l'attivazione del servizio di assegnazione di codici DOI rilasciati da DataCite grazie a una convenzione con la Conferenza dei Rettori delle Università Italiane (CRUI). Nonostante l'identificazione dei dati tramite DOI fosse disponibile, mancava tuttavia una strutturazione delle procedure scientifiche e amministrative per la validazione delle richieste dei colleghi che volevano condividere i propri dati. Emerse la necessità di dotarsi di una formalizzazione delle procedure da un lato, e dall'altro di avere una visione istituzionale dei principi che avrebbero dovuto guidare la condivisione dei dati per il prossimo futuro. Nel 2015 il Presidente INGV costituì un gruppo multidisciplinare composto da ricercatori, informatici e legali con il compito di redigere una Politica dei Dati istituzionale.

Una delle prime attività effettuate fu la mappatura del panorama normativo a livello nazionale, europeo e internazionale. Parallelamente si fece un censimento dei dati scientifici disponibili presso ciascuna sezione, in modo da avere contezza sia della varietà, sia della quantità di dati in gioco.

La redazione della Politica dei Dati durò dal 2015 al 2018 e si compone di tre documenti approvati dal Consiglio di Amministrazione dell'INGV. Un primo documento venne approvato nel 2016 e sancisce i Principi generali che si volevano perseguire. Un secondo documento venne approvato nel 2017 e tratta i principi dell'Open Access rivolti principal-

mente alle pubblicazioni scientifiche. Un terzo documento approvato nel 2018 descrive infine le modalità di attuazione di strumenti della Politica dei Dati che permettono la distinzione tra pubblicazioni di articoli scientifici e dati (Fig. 1).

Alcuni elementi presenti nella Politica dei Dati dell'INGV traggono spunto da indicazioni presenti in documenti redatti in vari ambiti della Ricerca e della Pubblica Amministrazione, sia a livello nazionale che internazionale. Ad esempio la costituzione di un "Ufficio Gestione Dati" che si occupi dell'attuazione della Politica dei Dati è stato ispirato dal "Gruppo di lavoro Open Data" descritto nelle Linee Guida Nazionali per la Valorizzazione del Patrimonio Informativo pubblico, a cura dell'Agenzia Digitale per l'Italia, la cui redazione è in continua evoluzione e aperta a nuovi contributi. L'adozione di un "Registro Dati" (<https://data.ingv.it>), che tra le funzioni svolte assegna anche identificativi univoci DOI secondo le specifiche DataCite, ha preso spunto da diversi cataloghi di metadati, come: quelli diffusi in ambito delle Direttive Europee INSPIRE che trattano la circolazione dei dati territoriali e dalle indicazioni fornite dalla Research Data Alliance. L'adozione di una catalogazione di dati basata su livelli di elaborazione e l'adozione di licenze di tipo Creative Commons è frutto di suggerimenti maturati nell'ambito dell'ERIC EPOS, l'European Plate Observing System e della sua Politica dei Dati.

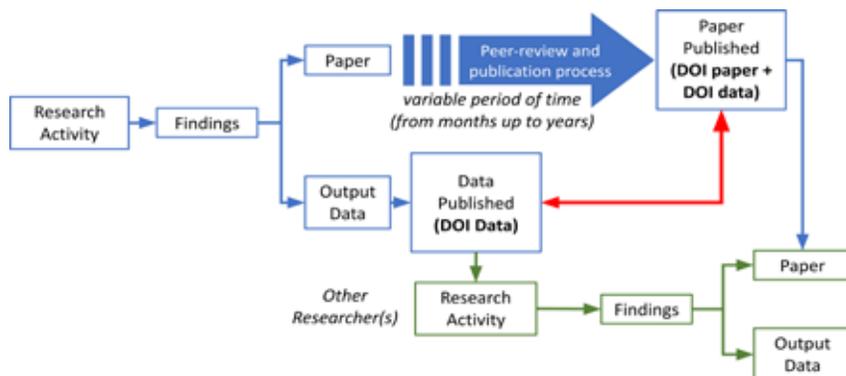


Fig. 1
Condivisione dei risultati della Ricerca tramite la separazione tra la pubblicazione di articoli e dati scientifici

La fase attuativa

L'attuazione della Politica dei Dati è iniziata nel novembre 2018 ed è in capo all'Ufficio Gestione Dati i cui componenti dedicano parte del proprio tempo a supporto di queste attività istituzionali. L'Ufficio è costituito da un Coordinatore, da un Referente per ciascuna delle tre aree scientifiche dell'istituto (i.e. Terremoti, Vulcani e Ambiente), e da personale informatico esperto nella gestione di infrastrutture digitali e di banche dati. Le attività dell'Ufficio sono pianificate su base annuale e vertono sulla progressiva adozione e diffusione dei principi FAIR (Findable, Accessible, Interoperable, Reusable) dei dati (Wilkinson et al. 2016), a cui si tende ad aggiungere il concetto di Reproducibility.

Lo strumento cardine per l'attuazione della Politica dei Dati è il Registro Dati, definibile in termini informatici come un "catalogo di metadati". Questo strumento è stato progettato e realizzato per interfacciarsi con i sistemi informativi dell'istituto come quello del personale e dei progetti di ricerca finanziati da altre organizzazioni. I metadati in esso contenuti

non seguono uno specifico standard, si tratta infatti di un'aggregazione di informazioni con standard diversi. Al momento sono stati sviluppati moduli di formattazione ed esportazione di metadati per supportare:

- lo standard DataCite per l'assegnazione di codici DOI;
- lo standard ISO 19115/19139 per la pubblicazione di metadati nel Repertorio Nazionale di Dati Territoriali;
- il profilo nazionale dei metadati DCAT-AP secondo le indicazioni dell'Agenzia per l'Italia Digitale.

Il Registro Dati mira a semplificare la "Findability" dei dati, in cui oltre alla descrizione dei dati condivisi, sono chiarite le regole di accesso e riutilizzabilità grazie all'apposizione di Licenze Creative Commons. Dal novembre 2020 il contenuto del Registro Dati è consultabile all'indirizzo <https://data.ingv.it> ed è basato sul software Open Source CKAN (<https://ckan.org/>).

Si mira a raggiungere l'interoperabilità dei dati e dei servizi di accesso ai dati tramite la progressiva adozione di servizi web che rispettino standard disciplinari come FDSN, International Federation of Digital Seismograph Networks, standard generici come OGC, Open Geospatial Consortium e le Direttive INSPIRE per i dati territoriali.

Le attività condotte per l'attuazione della Politica dei Dati dell'INGV vengono anche concretamente supportate dalla partecipazione attiva dell'Ente a iniziative quali la San Francisco Declaration on Research Assessment (DORA) sottoscritta nel 2018, che mira all'evoluzione dei sistemi di valutazione dei risultati della Ricerca, la partecipazione a iniziative di coordinamento nazionale come ICDI (Italian Computing and Data Infrastructure) coordinato dal GARR, o la partecipazione diretta alla costruzione dell'EOSC, l'European Open Science Cloud.

Riferimenti Bibliografici

Meadows, A. (2018), Enter once, reuse often, *Revista ORL*, 9(4), 301-302. <https://doi.org/10.14201/orl.18547>

Nielsen M., (2011), *Reinventing Discovery: The New Era of Networked Science*. Princeton University Press, 280 pp. <http://press.princeton.edu/titles/9517.html>

Wilkinson M. D., Dumontier M., Aalbersberg Ij. J., Appleton G., Axton M., Baak A., Blomberg N., Boiten J.-W., da Silva Santos L. B., Bourne P. E. et al. (2016), The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data*, 3(1), <https://doi.org/10.1038/sdata.2016.18>

Autori

Mario Locati mario.locati@ingv.it

Tecnologo all'INGV dal 2002. Si occupa della gestione di banche dati di terremoti e con questi partecipa a progetti nazionali ed europei. Promuove i dati aperti poiché si occupa di prodotti frutto dell'integrazione di dati da diverse fonti. Nel 2013 permette l'attivazione del servizio DataCite per l'assegnazione di codici DOI ai dati istituzionali. "Master in Management of Research, Innovation and Technology" presso il Politecnico di Milano nel 2017. Dal 2018 coordina l'Ufficio Gestione Dati.

Francesco Mariano Mele franco.mele@ingv.it

Laureato in fisica, è stato ricercatore del ING-INGV per oltre 36 anni, responsabile del CED ING, dei sistemi di calcolo real-time per il Servizio di Sorveglianza Sismica e del Bollettino Sismico. Si è occupato di calcolo di parametri ipocentrali, tomografia su scala locale e regionale, analisi di sequenze sismiche. È stato membro del Comitato Direttivo del Centro Allerta Tsunami dell'INGV dal 2017 al 2021 e referente per la Sismologia dell'Ufficio Gestione Dati dal 2018 all'aprile 2021.

Placido Montalto placido.montalto@ingv.it

Ricercatore presso l'Osservatorio Etneo, è laureato in ingegneria informatica con PhD in ingegneria dei sistemi. Responsabile della Sala Operativa e dei servizi di Information Technology dell'Osservatorio Etneo, si occupa della progettazione e realizzazione di sistemi per l'elaborazione e l'archiviazione dei dati, della progettazione e realizzazione di software per il processamento dei dati geofisici. È co-autore di oltre 40 pubblicazioni in ambito tecnico e scientifico.

Vincenzo Romano vincenzo.romano@ingv.it

PhD presso "University of Nottingham", Ingegnere Elettronico, Primo Tecnologo presso l'INGV. Esperienza ventennale nelle aree: Fisica Ionosferica e radiopropagazione, Space weather, GNSS, Data Management e Trasferimento tecnologico. Ha partecipato ad oltre 20 progetti di Ricerca e Sviluppo, nazionali e internazionali, spesso con ruoli di responsabilità. È stato fondatore e Direttore Generale del primo spin-off INGV. È co-autore di oltre 50 articoli su riviste peer-review internazionali.

Valentino Lauciani valentino.lauciani@ingv.it

Tecnico informatico presso INGV dal 2000. Partecipa alla progettazione e allo sviluppo di molti dei software e siti web utilizzati nella Sala di Sorveglianza Sismica di Roma. Si occupa anche di Hardware e nello specifico della gestione del Cluster VMWare. Sviluppa progetti su Azure con l'uso dei container e negli ultimi anni si dedica ai backend, progettando e sviluppando Database ed API nel rispetto dei vari standard esistenti. Il bello di questo lavoro? Ogni giorno si scoprono cose nuove ;-)

Roberto Vallone roberto.vallone@ingv.it

Laurea in Scienze della Terra. Tecnologo, presso INGV dal 2017. Gestisce la piattaforma tecnologica (hardware e software) che sta alla base del database delle faglie sismogenetiche sia a livello di area mediterranea a fini scientifici e di protezione civile, sia a livello europeo in ambito EPOS ERIC per la sismologia (<https://www.seisfaults.eu>) e la pericolosità da tsunami (<https://www.tsunamidata.org>). Dal 2019 fa parte dell'Ufficio Gestione Dati dell'Ente come referente informatico.

Stefano Cacciaguerra stefano.cacciaguerra@ingv.it

Laurea in Scienze dell'Informazione, Ph.D in Informatica e Master in "Management of Research, Innovation and Technology". Come Tecnologo INGV, si occupa del trasferimento tecnologico dell'innovazione nel campo dell'Informatica a favore della progettazione, realizzazione e gestione di nuove infrastrutture ICT per l'implementazione di applicazioni Geofisiche, Vulcanologiche e Oceanografiche per attuare a livello nazionale progetti europei come EOSC, EDI e EUROHPC. Responsabile del CED di Bologna.

EOSC and Research Infrastructure: Legal Challenges of HPC

Ludovica Paseri

University of Bologna

Abstract. The European Open Science projects are strictly related to the wider European Cloud Initiative. This Initiative also takes into account the development of the research infrastructures, and, in particular, the High Performance Computing for scientific research purposes is highly encouraged. The implementation of HPC platforms for the benefit of universities and research centres entails several advantages but also poses some legal challenges. The paper aims to describe the link between Open Science, the European Open Science Cloud (EOSC) and HPC, focusing, finally, on the legal challenges of the latter

Keywords. Open Science; European Polices; High Performance Computing; Scientific Research; European Cloud Initiative; EOSC

Introduction: The European Open Science Projects

The European institutions, and in particular the European Commission, are engaged in supporting scientific research by fostering Open Science projects. The Open Science paradigm advocates the openness of every phase of the research process, from data collection to publication of results (Fecher, Friesike, 2014), following the principle «as open as possible, as closed as necessary». The openness of research data promoted by the Open Science, in fact, does not aim at a blind and indiscriminate sharing of data. On the contrary, such openness is always the result of balancing the different interests involved (e.g., the protection of personal data; national security; or public interest).

On 23 November 2018, the European Open Science Cloud (EOSC hereinafter) was launched in Vienna: a trusted and federated environment that aims to coordinate the Open Science projects, implemented in recent years, to facilitate the sharing of scientific research both in terms of research data and in terms of research results, namely publications, for the benefit of 1.7 million European researchers and 70 million science and technology professionals, by breaking down geographical barriers, and boundaries traditionally imposed by different disciplines (Ayris, et al., 2016; Burgelman, 2021).

In the Communication of the European Commission entitled “A European Strategy for Data” (European Commission, 2020), the intention to support the Open Science is stressed and, in particular, the importance of EOSC is underlined. EOSC has been represented as one of the ten domain-specific data spaces for which the European institutions intend to develop specific actions and policies. An effective implementation of EOSC would be an example of digital sustainability, able to produce a concrete impact on the world of scien-

tific research and, more generally, on society.

In this context, characterised by the emergence of the Open Science paradigm, it is relevant to focus on the High Performance Computing (HPC, hereinafter), as a means by which foster scientific research (Section 1). Clarified this scenario, and the interplay between the Open Science projects and the HPC, attention will be drawn to the legal challenges which have arisen (Section 2). Finally, Section 3 will provide the conclusions of the analysis.

1. High Performance Computing for Scientific Research

EOSC should be considered as a part of a wider and more ambitious European project: the European Cloud Initiative (European Commission, 2016). This initiative, which aims to build a competitive data and knowledge economy in Europe, is based on two pillars: (i) EOSC; and (ii) the European Data Infrastructure (EDI, hereinafter). While EOSC is, as mentioned above, a trusted and federated ecosystem for sharing science in Europe, EDI is, instead, an infrastructure of highest quality, which will underpin EOSC, based on supercomputing capacity, fast connectivity and high-capacity cloud solutions (European Commission, 2016). An extremely relevant aspect of EDI architecture is represented by the High Performance Computing strategy (European Commission, 2012). HPC is a form of supercomputing that entails a complex processing system, involving multiple processors capable of providing high performance, previously unseen, such as real-time analysis of huge amounts of data, the so-called Big Data. The HPC strategy is the means by which the value of Big Data in Europe can be fully exploited, and by linking up with EOSC project, the main beneficiaries may become researchers, universities and research centres. The EDI structure takes into account the need of IT infrastructures by European researchers in order to take full advantage of the digital services available today.

The link between the development of Open Science projects (EOSC in particular), and the implementation of HPC platforms, as infrastructures to foster scientific research in Europe, is emphasised by the recent European Strategy for data: «The aim is to help common data and world class cloud infrastructures for the public good to emerge, enabling secure data storage and processing for the public sector and research institutions» (European Commission, 2020).

In order to foster HPC in Europe, the European High-Performance Computing Joint Undertaking (EuroHPC JU, hereinafter) has been created: it is a legal and funding entity which will empower pooling of EU and national resources in HPC. The EuroHPC JU project is explicitly related to the European Cloud Initiative «[...] in order to gather the necessary resources and capabilities [...] to close the chain from research and development to the delivery and operation of the exascale High Performance Computing system», as stated in the Council Regulation 2018/1488 of 28 September 2018. It is also claimed that «[...] a mechanism should be set up at Union level to combine and concentrate the provision of support to the establishment of a world-class European High Performance Computing infrastructure and for research and innovation in High Performance Computing by Member States, the Union and the private sector. This infrastructure should provide

access to the public sector users, users from industry, including small and medium- sized enterprises (SMEs), and users from academia, including the scientific communities of the emerging European Open Science Cloud» (Council Regulation, 2018).

2. HPC Legal Challenges

Providing exascale supercomputing capability to universities and research centres and boosting the European digital economy starting from research and innovation bring evident benefits. The novelty of these initiatives, however, also entails some previously unseen challenges. The major legal issues are summarised below in three macro categories: data protection; commercial trade secrets; ownership of data.

Data protection. The compliance with GDPR is complicated by the different layers of which an HPC platform is generally composed. The main issues are related to compliance with the principles of lawfulness, fairness and transparency, purpose limitation, data minimisation, accuracy, storage limitation, integrity, confidentiality, and accountability (Paseri, et al., 2021). Another interesting aspect is related to Article 22 GDPR, which requires that «The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her» (see, Pagallo, 2020). Although scientific research enjoys the derogations provided for in Article 89 GDPR, the complex context requires a more in-depth analysis. In addition, data protection issues in the specific field of HPC have been relatively little investigated so far. Considering this aspect and the typical data protection issues in the field of scientific research (e.g. discipline fragmentation, on the Italian case, see: Durante, Senor, 2019), the situation becomes more complex and deserves further investigation.

Commercial trade secrets. The topic becomes relevant when the research project is co-sponsored by private actors, who have economic interests: the participation of different kind of actors can lead to the situation in which researchers do not release research results until a patent has been obtained (Carroll, 2015). In this situation, the initial phase of the research project becomes crucial: it is in the setting of agreements between universities and private entities that the future of the project is defined. The hurdle is represented by the fact that very often research bodies have neither legal competence, nor time, nor bargaining power with private bodies. This lack of legal support can represent a limit for the subsequent sharing of research results.

Ownership of data. The data ownership raises two different levels of complexity: on the one hand there is a tension between those conducting the research and those running the platform (e.g., Sheehan, 2018); and on the other hand, in research projects involving private companies there may be a problem of ownership of the data processed and resulting by the research projects (Guibault, Wiebe, 2013).

3. Conclusions

In light of the analysis of the ecosystem designed by the European Union that involves Open Science projects, in particular the EOOSC and the HPC infrastructure, some legal

issues can be identified. This contribution aims to provide an overview of these challenges, which are: the protection of personal data involved, the management of trade secrets and the ownership of research data, in relation to the multitude of actors involved with different interests.

These issues need to be further investigated, considering that there are already a multitude of HPC platforms operating for scientific research, scattered across the EU. Usually, researchers and operators of these HPC platforms are dealing with the same legal issues, leading to a kind of general uncertainty. What is crucial to understand, however, is that very often this legal uncertainty does not stem from the lack of ad hoc legal provisions for the specific technology: more frequently the uncertainty concerns the application of existing disciplines. This uncertainty becomes a barrier to sustainable scientific research capable of exploiting the advantages of the most innovative technologies. Nevertheless, the commitment to the Open Science principles can transform this limit into an opportunity (Besançon, 2021), pursuing not only greater openness of science, but also greater collaboration between researchers and different fields of knowledge.

References

- Ayris P., et. al., (2016) Realising the European open science cloud, European Union.
- Besançon, L., et al. (2021) “Open science saves lives: lessons from the COVID-19 pandemic.” *BMC Medical Research Methodology* 21.1, 1-18.
- Burgelman, J.-C. (2021) “Politics and Open Science: How the European Open Science Cloud Became Reality (the Untold Story).” *Data Intelligence* 3.1, 5-19.
- Carroll, M., W., (2015) Sharing research data and intellectual property law: A primer, in “*PLoS biology*” 13.8.
- Council Regulation (EU) 2018/1488 of 28 September 2018, “Establishing the European High Performance Computing Joint Undertaking”, ST/10594/2018/INIT, in OJ L 252, 8.10.2018, p. 1–34.
- Durante, M., Senor, M., (2019) “Report on the harmonization of Italian Law with the enforcement of The GDPR.”, 89-98.
- European Commission, (2012) “High-Performance Computing: Europe's place in a Global Race”, COM/2012/45 final.
- European Commission, (2016) “European Cloud Initiative - Building a competitive data and knowledge economy in Europe”, COM/2016/178 final.
- European Commission, (2020) “A European Strategy for Data”, COM/2020/66 final.
- Fecher B., S. Friesike, (2014) Open science: one term, five schools of thought, in: “Opening science”, Springer, Cham, 2014. 17-47.
- Guibault, L., Wiebe, A. (2013) “Safe to be open: Study on the protection of research data and recommendations for access and usage”, Universitätsverlag Göttingen.
- Pagallo, U., (2020) “Algoritmi e conoscibilità.” *Rivista di filosofia del diritto* 9.1, 93-106.
- Paseri, L., Varrette, S., Bouvry, P., (2021) “Protection of Personal Data in High Performan-

ce Computing Platform for Scientific Research Purposes.” Annual Privacy Forum. Springer, Cham, 123-142.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

Sheehan, K. B., (2018) “Crowdsourcing research: data collection with Amazon’s Mechanical Turk.” Communication Monographs 85.1, 140-156.

Author



Ludovica Paseri ludovica.paseri2@unibo.it - ludovica.paseri@unito.it

Ludovica Paseri is a PhD candidate at LAST-JD Joint International Doctoral degree program in “Law, Science and Technology”, at the Alma Mater Studiorum, University of Bologna, in cotutelle with the University of Luxembourg, faculty of Computer Science. She graduated in Law at the University of Turin and attended a postgraduate Master in Data Protection Law, at the Department of Law, University of Turin. Ludovica has completed the internship as trainee lawyer in an Administrative Law firm and her main research interests are Cloud Computing, EU Open Science policies, scientific data governance, privacy and data protection law.

“Data steward” per i dati FAIR

Valentina Pasquale¹, Emma Lazzeri², Elena Giglia³

¹Istituto Italiano di Tecnologia, ²Consortium GARR, ³Università degli Studi di Torino

Abstract. I dati sono il fondamento di ogni ricerca scientifica e pertanto devono essere conservati e condivisi, quando possibile, per una scienza solida e riproducibile. Enti di ricerca in tutto il mondo si stanno dotando di centri di competenza in open science e data management per supportare i ricercatori nel rendere FAIR i dati della ricerca, attraverso l'introduzione dei data steward. Lo scopo di questo lavoro è inquadrare lo stato dell'arte della definizione della figura professionale del data steward e delle sue competenze, con particolare riferimento ai recenti e futuri sviluppi in Italia

Keywords. open science, FAIR data, research data management, data steward, EOSC

Introduzione: perché il data steward

I dati sono il fondamento di ogni ricerca scientifica. Pertanto devono essere conservati e, quando possibile, condivisi per una scienza solida e riproducibile. Inoltre, “l'utilizzo intelligente dei dati, ivi compreso il loro trattamento attraverso applicazioni di intelligenza artificiale, può trasformare tutti i settori dell'economia” (Direttiva EU 1024/2019). L'intelligenza artificiale ha però bisogno di dati “intelligenti”, che possono essere (ri)utilizzati solo se accompagnati da metadati e rappresentati per renderli comprensibili e sfruttabili dalle macchine.

A questa esigenza rispondono i principi FAIR (Findable, Accessible, Interoperable, Reusable o anche Fully Artificial Intelligence Ready) (Wilkinson et al., 2016). Enti di ricerca accademica e privata in tutto il mondo si stanno dotando di centri di competenza in data stewardship per i dati FAIR.

Nella visione della Commissione Europea e di Horizon Europe, i dati sono al centro della ricerca e dell'innovazione: per affrontare le grandi sfide del futuro (cambiamento climatico, economia sostenibile, etc.) occorre che siano disponibili in modo FAIR e aperto a ricercatori ed esperti. Per questo la Commissione ha investito in EOSC, European Open Science Cloud, un “web di dati e servizi interconnessi” i cui blocchi costitutivi saranno i dati FAIR. EOSC sarà uno “spazio di condivisione” in cui chi produce e chi ri-usa dati potrà collaborare, creando nuove idee e soluzioni.

È stato stimato che nei prossimi anni per realizzare e operare EOSC saranno necessari 500.000 data steward. In questo contesto, è necessario destinare il 5% dei fondi di ricerca a una corretta gestione dei dati, inclusa l'assunzione dei data steward (Mons, 2020). In questa prospettiva, istituti di ricerca e istruzione universitaria devono impegnarsi a formare questa nuova figura professionale in modo strutturato e ad integrarla nel supporto alla ricerca. Scopo di questo lavoro è inquadrare lo stato dell'arte della definizione della figura professionale del

data steward e delle sue competenze, con particolare riferimento ai recenti sviluppi in Italia.

1. Le competenze del data steward

Il data steward è una nuova figura professionale competente nella pianificazione e gestione responsabile dei dati della ricerca durante tutto il loro ciclo di vita. Diversamente dal data scientist o dal research software engineer, il ruolo del data steward è ancora poco conosciuto e caratterizzato. Stewardship può essere tradotto con “amministrazione, gestione”, ma perde la sua connotazione di “pianificazione e gestione responsabile ed etica delle risorse” (Wikipedia).

La professionalizzazione dei data steward e la creazione di centri di competenze sono individuati nella EOSC Strategic Research and Innovation Agenda quali fattori critici di successo per EOSC. Alcuni progetti quali FAIRsFAIR e EOSC Pillar, nonché EOSC Executive Board Skills and Training Working Group chiuso nel 2020, e la Research Data Alliance (RDA) stanno lavorando alla definizione di framework di competenze del data steward. RDA Interest Group for Professionalizing Data Stewardship ha diversi task group su integrazione della data stewardship nelle istituzioni, formazione e profili di carriera, definizione e certificazione delle competenze e sostenibilità finanziaria.

L’Olanda è la nazione che per prima ha inserito i data steward nelle istituzioni, a cominciare da TU Delft nel 2017 (Teperek and Plomp, 2019). Sempre in Olanda, nel National Programme Open Science è stato finanziato un progetto per la definizione delle competenze e della formazione dei data steward (Jetten et al., 2021) che ha prodotto una classificazione dei vari profili, ovvero Research, Policy e Infrastructure (Scholtens et al., 2019). Per ogni profilo sono stati definiti funzioni, competenze e risorse formative.

In Danimarca nel 2019 il National Coordination of Data Stewards Education (Wildgaard et al., 2020), con università ed industria, ha analizzato e sintetizzato lo stato della formazione sulla gestione dei dati di ricerca. Interessanti sono la definizione di data steward come figura-ponte fra le istituzioni, le infrastrutture e l’utente finale, e il riconoscimento della molteplicità di funzioni che il data steward può svolgere, a seconda del contesto in cui sia impiegato. Sono stati individuati 4 ruoli: agente di cambiamento, creatore, analista, amministratore. Questi 4 ruoli hanno un tratto comune: la competenza di base sui dati di dominio, alla quale vanno aggiunte le competenze più tecniche e trasversali. Il data steward è quindi una figura di professionalità elevata, che in primo luogo sa come trattare i dati di una specifica disciplina. Nella formazione dei data steward la parola chiave è flessibilità e un bilanciamento fra hard e soft skills.

Nell’ambito del progetto FAIRsFAIR è stata condotta un’analisi delle competenze richieste al data steward dal mondo del lavoro (Demchenko & Stoy, 2021). Nell’analisi viene evidenziata la molteplicità di queste, sintetizzata nella definizione di 7 estensioni al modello di riferimento per la Data Science, EDISON Data Science Framework, e in un elenco dettagliato di elementi da inserire nei curricula universitari.

2. La situazione in Italia

In Italia, come in molti Paesi, non esiste ancora un percorso formativo specifico e strut-

turato per il data steward. Tuttavia, molte sono le attività formative su Open Science e gestione dei dati che vengono proposte a varie figure professionali, tra cui i ricercatori. Negli ultimi anni progetti europei e iniziative, quali OpenAIRE, RDA, EOSC-Pillar, GO FAIR, OPERAS, hanno organizzato corsi ed eventi di formazione. Molti sono gli esempi di corsi trasversali offerti agli studenti di dottorato in varie università (Torino, Pisa, Scuola Normale Superiore, Genova, etc.).

In questo scenario uno dei principali attori in Italia è ICDI (Italian Computing and Data Infrastructure), il tavolo di lavoro creato dai rappresentanti di alcune tra le principali Infrastrutture di Ricerca e Digitali italiane per promuovere sinergie al fine di ottimizzare la partecipazione italiana alle attuali sfide europee, tra cui EOSC, la European Data Infrastructure e high-performance computing. Nel 2020 ICDI ha avviato la creazione del primo "Centro di Competenze" per EOSC, l'Open Science e la gestione dei dati FAIR (Lazzeri et al., 2021).

Il Centro di Competenze si pone come unico punto di riferimento nazionale per supporto, consulenza e formazione sui temi Open Science, FAIR data e EOSC, professionalizzazione dei data steward, ed infine servizi e strumenti per il FAIR by design tenendo conto delle specificità di settore. Un valore aggiunto è la creazione di percorsi formativi dedicati grazie alla rete di esperti di settore.

3. Conclusioni e prospettive future

L'Italia ha accumulato ritardo verso la realizzazione dell'Open Science e la formazione strutturata dei data steward. Una spinta positiva può venire dalla pubblicazione del Piano Nazionale per la Scienza Aperta, per ora solo anticipato nella sezione 6.2 del Programma Nazionale per la Ricerca 2021-2027. Il Piano prevede 4 assi: testi, dati della ricerca, valutazione della ricerca e coinvolgimento delle comunità. Nella sezione sui dati della ricerca si auspica la creazione di un data center nazionale che funga da centro di competenze.

Occorre un coordinamento nazionale sotto l'egida del Ministero dell'Università e della Ricerca, che promuova le pratiche di Open Science e crei sinergie sulla gestione dei dati con il Ministero della Salute per i dati sanitari e il Ministero dello Sviluppo Economico per la strategia sull'intelligenza artificiale.

Riferimenti bibliografici

- Demchenko, Y., & Stoy, L. (2021), Research Data Management and Data Stewardship Competences in University Curriculum, 2021 IEEE Global Engineering Education Conference (EDUCON), 1717–1726. <https://doi.org/10.1109/EDUCON46332.2021.9453956>
- Jetten, M., Grootveld, M., Mordant, A., Jansen, M., Bloemers, M., Miedema, M., & Gelder, C. W. G. van. (2021), Professionalising data stewardship in the Netherlands. Competences, training and education. Dutch roadmap towards national implementation of FAIR data stewardship, Zenodo. <https://doi.org/10.5281/zenodo.4623713>
- Lazzeri, E., Tanlongo, F., Pavone, G., Alpi, F., Ansuini, A., Bertazzon, E., Bonaccorsi, D., Cappelluti, F., Casati, S., Castelli, D., Cippitani, R., Colcelli, V., Costantini, A., Cozzini, S., Degl'Innocenti, E., Di Donato, F., Di Giorgio, S., Fava, I., Fiore, S., ... Zane, D. (2021), Competence Centre ICDI per Open Science, FAIR, ed EOSC - Mission, Strategia e piano

d'azione, Zenodo <https://doi.org/10.5281/zenodo.5071055>

Mons, B. (2020), Invest 5% of research funds in ensuring data are reusable. *Nature*, 578(7796), 491–491. <https://doi.org/10.1038/d41586-020-00505-7>

Scholtens, S., Jetten, M., Böhmer, J., Staiger, C., Slouwerhof, I., Geest, M. van der, & Gelder, C. W. G. van. (2019), Final report: Towards FAIR data steward as profession for the lifesciences. Report of a ZonMw funded collaborative approach built on existing expertise, Zenodo <https://doi.org/10.5281/zenodo.3474789>

Teperek, M., Plomp, E. (2019), The role and value of data stewards in Universities: a TU Delft case study on data stewardship, Zenodo <http://doi.org/10.5281/zenodo.2684278>

Wildgaard, L., Vlachos, E., Nondal, L., Larsen, A. V., & Svendsen, M. (2020), National Coordination of Data Steward Education in Denmark: Final report to the National Forum for Research Data Management (DM Forum), Zenodo <https://doi.org/10.5281/zenodo.3609516>

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016), The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>

Autrici

Valentina Pasquale valentina.pasquale@iit.it

Valentina lavora come "Research Data Management Specialist" all'Istituto Italiano di Tecnologia, dove coordina i servizi di supporto per la gestione dei dati della ricerca e la scienza aperta. Coordina la GO FAIR Data Stewardship Competence Centers Implementation Network. Prima di occuparsi di dati della ricerca, ha ottenuto un dottorato in "Humanoid Technologies" all'Università di Genova nel 2010 e lavorato per diversi anni nel campo delle Neuroscienze Computazionali.

Emma Lazzeri emma.lazzeri@garr.it

Emma lavora per Consorzio GARR alla definizione di strategie e alla formazione e informazione in tema di scienza aperta. Coordina la Task Force di ICDI per il Centro di Competenza Nazionale per l'Open Science, FAIR ed EOSC. È coordinatrice dei nodi italiani di OpenAIRE e della Research Data Alliance. Ha conseguito il dottorato di ricerca in Innovative Technologies-Telecommunications presso Scuola Superiore Sant'Anna e un Master in Ingegneria delle telecomunicazioni presso l'Università di Pisa.

Elena Giglia elena.giglia@unito.it

Elena è responsabile dell'Unità di Progetto Open Access all'Università di Torino. Svolge un ruolo attivo in gruppi di lavoro nazionali e internazionali sull'Open Science. Ha fatto parte presso il MIUR della Commissione che ha redatto il nuovo Piano Nazionale per la Scienza Aperta. È stata membro del gruppo di lavoro "Open Access e nuovi indicatori" presso ANVUR (2013-2014). Partecipa attivamente a convegni su Open Science e svolge una intensa attività di formazione, informazione e promozione.

Integrazione dei Principi FAIR nel processo di ricerca biomedica: la creazione del Registro della Regione Sardegna

Alessandro Sulis¹, Vittorio Meloni¹, Cecilia Mascia¹, Franco Cappai², Caterina G. Carboni², Ernesto D'Aloja³, Giorgio Fotia¹, Giuseppe Serra², Giovanni Sotgiu⁴, Paolo Uva¹, Francesca Frexia¹, Gianluigi Zanetti¹

¹CRS4: Centro di Ricerca, Sviluppo e Studi Superiori in Sardegna, ²Sardegna Ricerche, ³Università degli Studi di Cagliari, ⁴Università degli Studi di Sassari

Abstract. Il Registro per la Ricerca Biomedica della Regione Sardegna favorisce il riuso e la condivisione dei risultati della ricerca, sia da parte di esseri umani, sia di sistemi automatici (machine-actionability), in linea coi Principi FAIR. Questo lavoro descrive l'architettura del Registro, le scelte progettuali e implementative e gli sviluppi futuri in termini di completamento e messa in funzione del sistema. Il Registro è uno dei pilastri portanti del Programma I FAIR, iniziativa volta a promuovere le buone pratiche nella raccolta e gestione dei dati biomedici nell'ambito degli studi clinici indipendenti

Keywords. Ricerca biomedica, FAIR, Registro, Interoperabilità Semantica, FAIR Data Point, Molgenis

Introduzione

La difficoltà a impiegare in altri contesti i dataset creati durante lo svolgimento di un trial clinico (Vines et al. 2014), così come la bassa riproducibilità dei risultati ottenuti (Freedman et al. 2015), hanno generato un interesse via via crescente verso l'Open Science (Allen et al. 2019) e verso la formalizzazione di linee guida come i Principi FAIR (Wilkinson et al. 2016), che promuovono la creazione nel processo di ricerca di digital research objects (dati, algoritmi, strumenti, etc.) che siano Findable, Accessible, Interoperable e Re-Usable.

Il Registro per la Ricerca Biomedica della Regione Sardegna è uno strumento pensato per affrontare queste criticità supportando la condivisione secondo i Principi FAIR di informazioni descrittive (metadati) sui dataset creati nell'esecuzione di studi clinici indipendenti, compatibilmente con i principi etici, le disposizioni legali e la protezione della proprietà intellettuale. La realizzazione del Registro è uno dei pilastri fondamentali del Programma I FAIR (Cappai et al. 2019), illustrato in Figura 1, l'iniziativa di formazione e trasferimento tecnologico che ha come obiettivo quello di diffondere e promuovere buone pratiche nella raccolta, gestione e condivisione di dati biomedici, da un punto di vista tecnologico, etico e statistico.

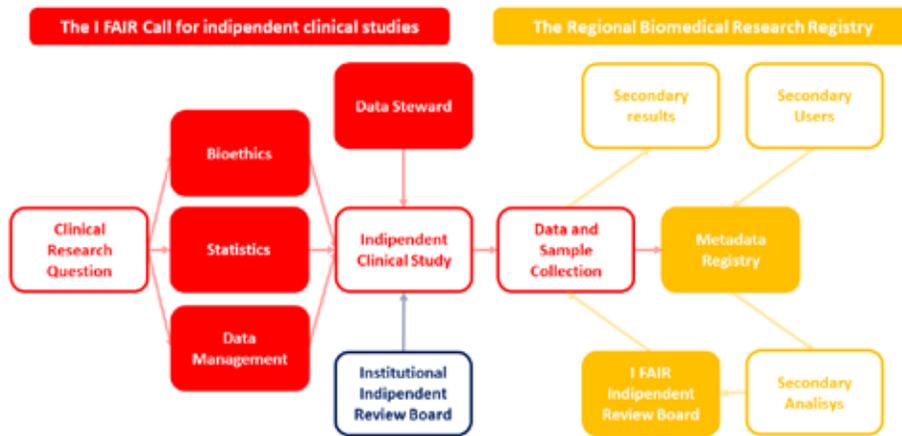


Fig. 1
Il programma
IFAIR della
Regione
Sardegna

1. Il Registro

Il Registro è stato progettato e realizzato per essere una risorsa FAIR, ovvero una collezione di “oggetti” caratterizzata da contenuti e funzionalità accessibili e utilizzabili direttamente da esseri umani e sistemi automatici (machine-actionability). L’implementazione segue le principali best-practices per il processo di FAIRificazione (Jacobsen et al. 2020), attuate secondo le fasi rappresentate nella Figura 2.

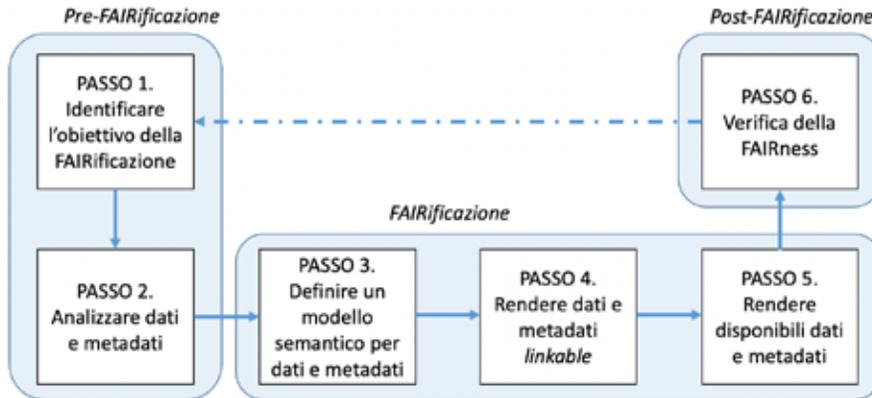


Fig. 2
Workflow di
FAIRificazione,
(riadattato da
Jacobsen, 2020)

Il primo passo del workflow è la definizione dell’obiettivo del processo, che coincide con l’obiettivo operativo del Registro: la condivisione di informazioni aggregate e parametri descrittivi sugli studi clinici (metadati). Una prima selezione di metadati comuni ai vari studi è quindi stata analizzata (Passo 2) per definire la strategia per la fase di FAIRificazione vera e propria, che prevede: la definizione del modello dei concetti (modello semantico) descritti da dati e metadati (Passo 3), la loro trasformazione in un formato machine-readable (Passo 4), la realizzazione di uno strato di persistenza e condivisione (Passo 5). Il Registro implementa questi passi seguendo le specifiche FAIR Data Point

(FDP) (Da Silva Santos et al. 2016), che mettono a disposizione un Semantic Data Model, basato sull'ontologia Data Catalog Vocabulary (DCAT) (Maali et al. 2014), insieme a un set di API per popolare e interrogare un registro di metadati nel rispetto dei Principi FAIR. I metadati selezionati per gli studi sono stati quindi mappati, ove possibile, sul modello semantico fornito dalle specifiche del FDP. Per i metadati più specifici, come ad esempio “area terapeutica” e “tipologia di dati raccolti”, per i quali non è stata trovata una corrispondenza diretta con i concetti del modello FDP, il modello stesso è stato esteso utilizzando la Semantic Science Integrated Ontology (SIO) (Dumontier et al. 2014). I relativi metadati sono stati quindi valorizzati mediante le principali ontologie in uso nel dominio clinico (e.g., ICD e MESH), in modo da definire ogni valore senza ambiguità, a supporto dell'interoperabilità semantica.

L'architettura e l'implementazione del Registro sono basate su Molgenis (van der Velde et al. 2019), un'applicazione open source per la raccolta, la gestione e la condivisione di dataset di ricerca. Molgenis implementa nativamente le specifiche del FDP e permette la creazione di interfacce grafiche per la consultazione e la gestione dei dataset. Nel contesto del Registro è stato esteso il modello semantico del FDP di Molgenis ed è stata creata un'interfaccia grafica per la ricerca e l'accesso ai metadati degli studi. Per quanto riguarda la fase di Post-FAIRificazione (Passo 6), sono stati predisposti alcuni test per valutare l'effettiva copertura del sistema rispetto ai Principi FAIR, sulla base di specifiche metriche (Wilkinson et al. 2019).

1.2 Stato del prototipo

La versione attuale del Registro prevede un'istanza di Molgenis apposita, all'interno della quale il modello FDP esteso è stato popolato con un primo insieme di metadati generici relativi ai 18 studi facenti parte del Programma I FAIR. Il prototipo è dotato di un'interfaccia utente per la ricerca degli studi sulla base di metadati di diverso tipo e per la consultazione delle informazioni di dettaglio relative a ciascuno studio. Per ogni parametro, la ricerca è ristretta ai valori possibili dello specifico concetto ontologico associato: ad esempio, per il parametro “area terapeutica” sarà possibile effettuare la ricerca solo sulla base dei valori relativi dell'ontologia MESH.

2. Conclusioni

Il Registro per la Ricerca Biomedica della Regione Sardegna è uno strumento realizzato secondo i Principi FAIR e finalizzato a raccogliere un insieme di metadati descrittivi relativi a studi clinici indipendenti, per facilitare la condivisione e il riuso dei dati raccolti. Il Registro è in fase di popolamento con i metadati degli studi selezionati dal Programma I FAIR, mentre in seguito verranno compresi anche i riferimenti ad altri studi che vorranno aderire.

Sviluppi futuri prevedono l'estensione del modello semantico del Registro, ad esempio per includere concetti che possano descrivere, quando disponibili, la provenance dei dati generati nel processo di ricerca, al fine di supportare un miglioramento della riproducibilità dei risultati. Il Registro può essere facilmente riadattato per raccogliere informazioni

descrittive riguardo a tipi diversi di studi di ricerca. Per favorire un riuso in altri contesti che richiedano la condivisione di informazioni su studi clinici secondo i Principi FAIR, il modello semantico è stato pubblicato (ifair-reg, 2021) e al termine dello sviluppo il codice verrà rilasciato in open source. L'intero approccio proposto dal Programma I FAIR, che combina le soluzioni tecnologiche con formazione e supporto multidisciplinare in un percorso verso una gestione e condivisione più "FAIR" dei risultati della ricerca, si candida ad essere considerato un modello di riferimento, in quanto l'analisi effettuata durante la preparazione del piano di attività non ha evidenziato la presenza di iniziative simili nel contesto di studi clinici indipendenti.

Ringraziamenti

I contributi degli autori sono realizzati nell'ambito del Programma I FAIR (finanziamento da Sardegna Ricerche, nell'ambito del POR FESR 2014/2020, Progetto FAIR DATA), del Progetto DIFRA (finanziamento dalla Regione Autonoma della Sardegna) e del Progetto EJP-RD (finanziamento H2020, Grant n. 825575).

Riferimenti bibliografici

- C. Allen, D. M. Mehler (2019), Open science challenges, benefits and tips in early career and beyond, *PLoS biology* 17 (5).
- F. Cappai, C. G. Carboni, E. D'Aloja, G. Fotia, F. Frexia, G. Serra, G. Sotgiu, P. Uva, G. Zannetti (2019), I FAIR Program: the Sardinian way to support and fund independent clinical studies that want to be Findable Accessible Interoperable Reusable, in: *The Ecosystem of Evidence Conference - Abstract book*, GIMBE, Evidence for Health, p. 15.
- L. B. Da Silva Santos, M. D. Wilkinson, et al. (2016), FAIR data points supporting big data interoperability, *Enterprise Interoperability in the Digitized and Networked Factory of the Future*. ISTE, London, pp. 270–279.
- M. Dumontier, C. J. Baker, et al. (2014), The SemanticScience Integrated Ontology (SIO) for biomedical research and knowledge discovery, *Journal of biomedical semantics* 5 (1), pp. 1–11.
- L. P. Freedman, I. M. Cockburn, T. S. Simcoe (2015), The economics of reproducibility in preclinical research, *PLoS Biol* 13 (6) e1002165.
- A. Jacobsen, R. Kaliyaperumal, et al. (2020), A generic workflow for the data FAIRification process, *Data Intelligence* 2 (1-2), pp. 56–65.
- ifair-reg (2021). URL: <https://github.com/crs4/ifair-reg>.
- F. Maali, J. Erickson, P. Archer (2014), Data catalog vocabulary (DCAT), *W3C recommendation* 16.
- K. J. van der Velde, F. Imhann, et al. (2019), MOLGENIS research: advanced bioinformatics data software for non-bioinformaticians, *Bioinformatics* 35 (6), pp. 1076–1078.
- T. H. Vines, A. Y. Albert, et al. (2014), The availability of research data declines rapidly with article age, *Current biology* 24 (1), pp. 94–97.
- M. D. Wilkinson, M. Dumontier, et al. (2016), The FAIR guiding principles for scientific

data management and stewardship, *Scientific data* 3 (1), pp. 1–9.

M. D. Wilkinson, M. Dumontier, et al. (2019), Evaluating FAIR maturity through a scalable, automated, community-governed framework, *Scientific data* 6 (1), pp. 1–12.

Autori



Alessandro Sulis alessandro.sulis@crs4.it

Alessandro Sulis è un ricercatore del gruppo Digital Health del CRS4 dal 2007. Ha un'esperienza decennale nell'interoperabilità tra sistemi clinici secondo i principali standard e best practices e collabora attivamente con gli organismi internazionali di riferimento (IHE e HL7 International). I suoi attuali interessi di ricerca comprendono lo studio e la promozione dei Principi FAIR, nel contesto di progetti legati alla FAIRificazione di dati e metadati in ambito clinico.

Vittorio Meloni vittorio.meloni@crs4.it

Dalla laurea in Informatica ottenuta presso l'Università degli Studi di Cagliari nel 2008, si occupa di ricerca e sviluppo nel campo dell'Informatica Clinica e delle Applicazioni Biomediche. Lavora dal 2010 nel gruppo Digital Health del CRS4, nel quale è impegnato in progettazione e sviluppo di software focalizzati su integrazione di domini clinici tramite standard e linee guida internazionali, condivisione di dati della ricerca secondo i Principi FAIR e telemedicina in tempo reale.



Cecilia Mascia cecilia.mascia@crs4.it

Cecilia Mascia, ingegnere biomedico, dal 2014 fa parte del gruppo Digital Health del CRS4, occupandosi principalmente dello sviluppo di modelli computabili di dati clinici mediante standard e ontologie internazionali (e.g., openEHR, HL7, OMOP), orientati all'interoperabilità semantica e al riuso dei dati in ottica FAIR. Attualmente collabora con il consorzio BBMRI-ERIC alla stesura di un proposal per lo Standard ISO 23494 dedicato alla cattura della provenance di dati e campioni biologici.

Franco Cappai cappai@sardegna ricerche.it

Franco Cappai fa parte dell'Unità di Supporto alla Ricerca Biomedica di Sardegna Ricerche dove coordina il Programma I FAIR. Ha un'esperienza ventennale nel trasferimento tecnologico e nella valorizzazione della ricerca con specializzazione in ambito biomedico. Attualmente partecipa al gruppo di lavoro ELSI di BBMRI-IT sul "Biobancaggio traslazionale" ed è patient expert reviewer per BMJ e NIHR.

Caterina G. Carboni carboni@sardegna ricerche.it

Caterina G. Carboni, laurea in Chimica e Tecnologie Farmaceutiche, dopo essersi occupata della gestione tecnico scientifica dei progetti in sviluppo preclinico per una società di sviluppo di nuovi agenti diagnostici e terapeutici, dal 2016 collabora con Sardegna Ricerche nell'organizzazione della Filiera Biomed occupandosi prevalentemente dell'introduzione dei Principi FAIR nella ricerca clinica spontanea ed indipendente e dell'istituzione del Registro della Ricerca Biomedica in Sardegna.

Ernesto D'Aloja ernestodalaja@gmail.com

Ernesto D'Aloja è Professore Ordinario di Medicina Legale nella Facoltà di Medicina e Chirurgia dell'Università di Cagliari e Direttore della Scuola di Specializzazione in Medicina Legale della stessa Università. È autore di oltre 140 pubblicazioni (I.F.>114) su riviste nazionali ed internazionali in tema di genetica forense, responsabilità professionale medica, patologia e tossicologia forense, e bioetica.

Giorgio Fotia giorgio.fotia@crs4.it

Giorgio Fotia è il direttore del Programma di Bioinformatica del CRS4. Il suoi attuali interessi di ricerca si concentrano sui metodi numerici ad alte prestazioni per la simulazione di problemi di biologia computazionale, sull'integrazione e sullo sviluppo di tecnologie abilitanti per l'analisi di dati biomedicali, e sulla loro applicazione alla soluzione di problemi data-intensive delle bioscienze. Attualmente è vicepresidente della Società Italiana di Matematica Industriale e Applicata ed Editor in Chief di Communications in Applied and Industrial Mathematics.

Giuseppe Serra giuseppe.serra@sardegnaricerche.it

Giuseppe Serra è responsabile del settore Startup e sviluppo del Parco scientifico e tecnologico dell'agenzia regionale Sardegna Ricerche. Al settore fanno capo le attività di supporto alla creazione e sviluppo di startup innovative e spin off dalla ricerca, la partecipazione a diversi programmi UE in materia di innovazione e trasferimento tecnologico e la gestione di programmi della Piattaforma biomed del parco e dell'Unità di supporto alla ricerca biomedica. Nel corso della sua esperienza ha ricoperto il ruolo di amministratore presso alcuni centri di ricerca regionali.

Giovanni Sotgiu gsotgiu@uniss.it

Giovanni Sotgiu è professore ordinario presso l'Università degli Studi di Sassari nelle tematiche di statistica medica, metodologia statistica applicata in ambito biomedico e clinico, organizzazione sanitaria, management e gestione sanitaria, informatica. Ha svolto e svolge ruolo in progetti di ricerca nazionali ed internazionali. Membro di diverse società scientifiche nazionali ed internazionali e di board editoriali di riviste internazionali.

Paolo Uva paolouva@gaslini.org

Paolo Uva, precedentemente responsabile della Piattaforma di Next Generation Sequencing del CRS4 è responsabile del gruppo di Bioinformatica Clinica presso l'IRCCS Gaslini (Genova) da maggio 2020. Le sue attività di ricerca sono centrate sull'analisi dei dati generati con la tecnologia del sequenziamento massivo e l'integrazione con altre tecnologie omiche. Le principali applicazioni riguardano lo studio delle basi genetiche delle malattie ereditarie.



Francesca Frexia francesca.frexia@crs4.it

Francesca guida il Programma di Ricerca Digital Health del CRS4. I suoi presenti interessi di ricerca si concentrano su modellazione di dati e processi biomedici e applicazione dei Principi FAIR, con un'attenzione particolare agli aspetti di interoperabilità e tracciabilità. Sta attualmente lavorando sulla modellazione dati con il formalismo openEHR e sulla stesura di linee guida e standard per la tracciabilità del campione (IHE SET Profile) e la provenance nelle biotecnologie (ISO 23494).



Gianluigi Zanetti gianluigi.zanetti@crs4.it

Gianluigi Zanetti è stato il Direttore del settore Data-intensive Computing del CRS4. Le sue attività di ricerca hanno toccato molte aree della scienza computazionale, tra cui la meccanica statistica computazionale e la simulazione del flusso sanguigno. Negli ultimi anni, si è dedicato in particolare allo sviluppo di strategie e infrastrutture in grado di garantire, in modo completamente tracciabile e riproducibile, l'elaborazione e l'analisi di grandi quantità di dati eterogenei e complessi.

EOSC-Pillar: dall'analisi giuridica alla definizione di policy per la Scienza Aperta

Sara Di Giorgio¹, Nadina Foggetti², Caterina Sganga³

¹Consortium GARR, ²INFN, ³Scuola Superiore Sant'Anna

1. Introduzione

Il progetto europeo EOSC-Pillar coordina le iniziative nazionali di open science in Austria, Belgio, Francia, Germania e Italia per facilitare lo sviluppo e l'adozione di politiche comuni e semplificare il processo di adesione all'European Open Science Cloud per i fornitori di servizi e le comunità di utenti. La definizione di policy comuni per la condivisione, l'accesso e il riuso dei risultati e degli strumenti digitali della ricerca è tra gli obiettivi principali del progetto. Tuttavia è necessario che queste policy siano basate sul quadro giuridico di riferimento degli Stati coinvolti, nonché delle norme internazionali ed europee applicabili in materia di tutela della proprietà intellettuale e di tutela dei dati personali e non personali.

La condivisione, l'accesso transfrontaliero e l'analisi dei dati della ricerca resi disponibili attraverso EOSC sono fortemente condizionati dalla normativa europea e nazionale e dalla comprensione che i ricercatori hanno del complesso quadro giuridico. Il presente lavoro ha l'obiettivo di presentare i primi risultati ottenuti dall'analisi del quadro normativo di riferimento (soft law e normativa vincolante), nonché della giurisprudenza prevalente in materia di diritto di proprietà intellettuale e protezione dei dati, mettendo in luce le differenze e i punti comuni, utili per sviluppare politiche europee per la condivisione dei risultati della ricerca. Sulla base dei risultati ottenuti dallo studio comparativo sono state elaborate delle linee guida che hanno l'obiettivo di agevolare il lavoro dei ricercatori durante l'elaborazione, la pubblicazione, la condivisione e l'integrazione dei dati di ricerca.

Lo scopo delle linee guida è quello di promuovere l'attuazione dei principi FAIR al di là del loro campo di applicazione originale, e di porre le condizioni per l'effettiva realizzazione di politiche di Open Data e Open Science. Facendo leva sulle flessibilità normative e tenendo conto delle legittime restrizioni all'accesso, le linee guida messe a punto da EOSC-Pillar mirano a (a) guidare i ricercatori nella gestione dei dati di ricerca e, più in generale, dei risultati della ricerca, e (b) promuovere le migliori pratiche per raggiungere l'accessibilità, la ricercabilità e l'interoperabilità dei dati di ricerca favorendo l'accesso aperto delle risorse e il loro riutilizzo, rimuovendo restrizioni inutili e gli ostacoli che possono rappresentare una barriera alla diffusione della Scienza Aperta, facilitando le convergenze delle soluzioni nazionali e mettendo in evidenza gli aspetti già sottoposti ad armonizzazione, grazie all'intervento normativo attuati in questi settori dall'UE.

Oltre alle Linee Guida, sono state elaborate delle Raccomandazioni rivolte ai respon-

sabili politici per suggerire delle riforme legislative e non legislative, a livello nazionale ed europeo, per colmare le lacune del quadro normativo analizzato, risolvere la frammentazione normativa e le divergenze tra gli Stati membri e aprire la strada verso un ambiente di ricerca efficace, aperto e inclusivo nell'EOSC e nell'UE.

2. L'elaborazione delle linee guida

EOSC-Pillar, nell'elaborazione di linee guida condivise per la Scienza Aperta, coglie le sfide proposte nell'ambito dell'agenda 2030 (<https://unric.org/it/wp-content/uploads/sites/3/2019/11/Agenda-2030-Onu-italia.pdf>) in merito alla sostenibilità, nonché del programma Nazionale per la ricerca 2021-2027 in relazione all'Open Science (<https://www.mur.gov.it/sites/default/files/2021-01/Pnr2021-27.pdf>). L'accessibilità, la riusabilità dei dati e l'interoperabilità dei dati rappresentano un aspetto cruciale nell'affermazione di una società tecnologicamente sostenibile ed aperta by design ed, in questa prospettiva, la definizione di policy condivise per la scienza aperta, rappresenta un obiettivo fondamentale (Bennato, 2020).

Partendo dalla normativa in materia di tutela dei dati personali e non personali nonché delle norme regolanti a livello europeo e transazione le proprietà intellettuale, l'attività di ricerca è stata volta ad individuare, mediante un gap analysis, i maggiori ostacoli che si frappongono alla definizione di un modello giuridico condiviso. Tra questi un ruolo importante assumono i conflitti di legge applicabile ad esempio in materia di tutela dei dati personali, la frammentazione della normativa settoriale, disciplinata dai codici di condotta regolanti alcuni settori specifici come i dati sanitari e genetici (M. Nanni, 2020). La disciplina del diritto dell'UE è intervenuta al fine di mitigare i rischi connessi ad una applicazione a geometria variabile delle discipline. In materia di tutela della proprietà intellettuale si registra ancora, dal punto di vista del diritto interno, la mancanza di standard condivisi a livello tecnologico e la presenza di un elenco relativo alle eccezioni e limitazioni alla disciplina che si configura come chiuso, non facilmente flessibile e diversificato all'interno dei singoli stati.

Un assunto fondamentale che ha guidato l'elaborazione delle linee guida è la necessità di garantire il principio dell' "open by design" mediante un'interpolazione lineare tra regole per la tutela dei dati e il rispetto delle norme in materia di tutela delle proprietà intellettuale. In questa definizione un ruolo centrale assume indubbiamente la definizione di data management plan chiari e precisi che permettano di raggiungere gli obiettivi relativi all'Open Science e all'Open Access, pur garantendo l'applicazione del legal framework di riferimento.

Le linee guida mirano a supportare i ricercatori in tutto il percorso di realizzazione del progetto di ricerca sin dalla fase della progettazione, infatti contengono delle specifiche regole che tengono conto non solo della normativa, ma anche delle best practices, nonché della disciplina contrattuale, sollecitando la definizione sin dalla fase di progettazione degli obiettivi di exploitation dei risultati. Le linee guida elaborate invitano, inoltre, sin da dalla definizione del proposal, a coinvolgere il team per il trasferimento tecnologico, ovvero a prevedere già la necessità di una valutazione di impatto per il trattamento dei

dati ai sensi del GDPR, o ancora considerare l'utilizzo di standard tecnologici e giuridici che consentano l'implementazione dei principi FAIR by design.

Nella fase di implementazione dell'attività di ricerca, le linee guida offrono supporto ai ricercatori per la definizione del piano di gestione dei dati alla luce delle norme relative alla proprietà intellettuale, della valutazione della flessibilità delle IPR, quali passaggi da seguire per garantire un'applicazione conforme ai principi del GDPR. In questo specifico ambito un'importanza cruciale assume la definizione della tipologia dei dati, delle modalità di utilizzo degli stessi al fine di garantire conformità al data management plan e la riutilizzabilità dei dati e dei risultati della ricerca. Nella fase finale del progetto, le linee guida sollecitano all'adozione di licenze d'uso dei dati, al fine ultimo di garantire il riuso dei dati e la concreta applicazione dei principi FAIR/OS/OA.

3. Policy Recommendations EOSC_Pillar Project

Le linee guida sopperiscono a gran parte dei gap normativi che abbiamo individuato, tuttavia alcuni aspetti di problematicità permangono rispetto all'applicazione del quadro normativo. In materia di tutela della proprietà intellettuale, le raccomandazioni rivolte ai policy maker sollecitano in primo luogo a portare avanti e migliorare l'approccio ad eccezioni e limitazioni già adottato dalla Direttiva 2019/790 sul mercato unico digitale, allo scopo di garantire un adeguato livello di armonizzazione in questo settore (P. Guarda, *Il regime giuridico dei dati della ricerca scientifica*, 2021), attraverso la definizione delle stesse quali obbligatorie ed inderogabili via contratto ove necessario a garantire l'adeguato bilanciamento tra diritto d'autore e libertà di ricerca. Sul piano della disciplina concernente le licenze, si evidenzia la necessità di un'innovazione normativa che miri ad agevolare la definizione di accordi di licenza software adeguati al settore scientifico. L'obiettivo è quello di garantire l'applicazione by default del diritto alla seconda pubblicazione dei risultati della ricerca e dell'applicazione del paradigma dell'open repository. Si rende necessario, in tal senso, un intervento armonizzatore nella materia della contrattualistica di diritto d'autore, al fine di garantire maggiore certezza del diritto ed uniformità nel diritto applicabile alle licenze, e dunque minori ostacoli alla circolazione transfrontaliera dei prodotti della ricerca. Le raccomandazioni mirano a sollecitare l'adozione a livello internazionale di un elenco unico, flessibile, armonizzato e obbligatorio di limiti ed eccezioni alle IPR che garantisca una maggiore chiarezza nella diffusione della conoscenza e il diritto alla scienza, mediante la diffusione dei risultati della ricerca e l'accessibilità e disponibilità delle pubblicazioni scientifiche.

In materia di Data protection, le raccomandazioni sollecitano il legislatore europeo e gli Stati della Comunità internazionale ad operare un bilanciamento tra standardizzazione, Open Access e Open Science, mediante l'introduzione di norme sempre più uniformi. Avuto specifico riguardo al trattamento dei dati particolari che richiedono una specifica regolamentazione, emerge la necessità di un processo di armonizzazione della disciplina concernente le procedure pseudonimizzazione necessarie al fine di garantire un trattamento dei dati genetici e/o genomici e relativi alla salute conforme alla disciplina dettata

a livello europeo e nazionale. Al contempo nelle linee guida si auspica una riforma normativa e regolamentare volta a garantire una semplificazione delle procedure di pseudonimizzazione per agevolare l'apertura dei dati stessi. Un altro aspetto di cruciale importanza riguarda la governance dei dati. Si chiede infatti al legislatore di definire in modo chiaro e preciso ruoli e responsabilità all'interno degli enti di ricerca nella gestione dei dati al fine di garantire l'applicazione dei principi OS OA.

4. Conclusioni

EOSC-Pillar nell'ambito dello studio condotto sul legal framework mira a creare delle linee guida che possano porre le basi per definire buone prassi da utilizzare a livello internazionale e nazionale che garantiscano da un lato la tutela dei diritti fondamentali e dall'altro garantiscano l'applicazione dei principi della Scienza Aperta.

La ricerca condotta integra un approccio dal basso verso l'alto raccogliendo i requisiti e le esigenze che emergono dalle diverse comunità scientifiche e uno dall'alto verso il basso, armonizzando le strategie europee e nazionali e fornendo un quadro chiaro di raccomandazioni e linee guida per far sì che le politiche nazionali che si vanno sviluppando, si traducano in una realtà operativa.

Il gruppo di ricerca di EOSC Pillar è impegnato nella raccolta di ulteriori requisiti e esigenze dei ricercatori e nell'aggiornamento dei riferimenti normativi a seguito dell'adozione della nuova disciplina a livello nazionale ed europeo in questo settore.

Bibliografia

Assemblea generale delle nazioni unite Settantesima sessione punti dell'agenda 15 e 116 Distr.: Generale 21 ottobre 2015 Risoluzione adottata dall'Assemblea Generale il 25 settembre 2015 Trasformare il

<https://unric.org/it/wp-content/uploads/sites/3/2019/11/Agenda-2030-Onu-italia.pdf>

CIPE il Programma nazionale per la ricerca 2021-2027 approvato il 15 dicembre 2020, <https://www.mur.gov.it/sites/default/files/2021-01/Pnr2021-27.pdf>

M. Nanni et. Al, Give more data, awareness and control to individual citizens, and they will help COVID-19 containment Transactions on data privacy, 2020, 13, 61 ff.

P. Guarda, Il regime giuridico dei dati della ricerca scientifica, Trento, 2021

Autori



Sara Di Giorgio sara.digiorgio@garr.it

is the Open Science Policy & International relation officer at GARR. She supports ICDI (Italian Computing and Data Infrastructure) and is involved in the EOSC-Pillar project.

Since 2003 consultant of the Ministry of Culture as coordinator of CulturalItalia. She is project manager of INDICES, she has been involved in several European projects for the development of research infrastructures in the field of cultural heritage and digital humanities, with particular focus on FAIR data production and management, Open Science Policies and in federating services through NREN e-infrastructures.

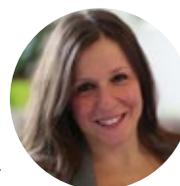


Nadina Foggetti nadina.foggetti@ba.infn.it

Fellow of INFN Bari. Degree in Law, Master in EU and transnational law -UniTN, Advanced course in Data Protection and Data Governance at and in Coding for Lawyers and Legal Tech -Unimi. PhD in international and EU law at the University of Bari A. Moro. Lawyer and mediator. Member of SIDI. Since 2005 she has collaborated in important national and international research projects in the fields of cybercrime, cybersecurity, data privacy, ITC law and cloud computing. In the EOSC-Pillar Project she contributes to WP 6 Task 6.6 Legal and ethical requirements for the storage and processing of health data and WP 4 Task 4.1 Policy and Legal Framework

Caterina Sganga caterina.sganga@santannapisa.it

joined Sant'Anna as an Associate Professor of Comparative Private Law in 2018. Assistant and later Associate Professor of Law at the Department of Legal Studies and Department of Economics and Business of Central European University (2012-2018). Ph.D. in Comparative Private Law, Sant'Anna (2011); LL.M. Yale Law School (2009), LL.B. and J.D. - University of Pisa (2004, 2006). Member elected of the European Copyright Society, fellow at the European Law Institute, member of several international IPR associations (EPIP, ATRIP, ALPS). From 2020, coordinator of the H2020 project reCreating Europe. She is involved in the EOSC-Pillar project.



Didattica digitale ed inclusione: il progetto Suoniamo

Maria Claudia Buzzi¹, Marina Buzzi¹, Marco Maugeri¹, Gabriella Paolini²,

Alessandra Sbragia², Caterina Senette¹, Amaury Trujillo¹

¹IIT-CNR, ²Consortium GARR

Abstract. L'autismo è un disturbo del neurosviluppo che comporta deficit nella comunicazione e interazione sociale, e deficit nel comportamento di immaginazione (attività e interessi ristretti, comportamenti ripetitivi e stereotipati). Molti studi confermano l'effetto positivo della musica sulla comunicazione, socializzazione e sul comportamento delle persone con autismo, ma per esse può essere difficile imparare il linguaggio musicale o a suonare uno strumento. Il progetto Suoniamo vuole sfruttare l'attrazione dei ragazzi con autismo nei confronti delle tecnologie per favorire l'apprendimento della musica. A questo scopo è stata sviluppata un'applicazione web che implementa un percorso didattico su una tastiera virtuale di pianoforte accessibile e personalizzabile. Ulteriori obiettivi sono favorire l'autonomia e l'autostima dell'alunno/a con autismo e agevolare l'inclusione scolastica, portando tutti i ragazzi della classe a collaborare insieme per l'esecuzione di semplici brani musicali

Keywords. Autismo, Educazione musicale, Tecnologie web, Inclusione, Didattica digitale

Introduzione

Il Manuale diagnostico e statistico dei disturbi mentali (APA, 2013) definisce il disturbo dello spettro autistico (ASD) come un disordine dello sviluppo neurologico con la presenza di ritardi e/o anomalie che compaiono prima dei 3 anni di età in almeno una delle aree: comunicazione sociale, interessi e/o comportamenti ristretti e ripetitivi.

L'ASD è una sindrome che impatta fortemente sulla vita di ogni giorno delle persone che ne sono affette e compromette la loro capacità di interpretare il mondo che le circonda. I sintomi variano significativamente da persona a persona: possono esserci deficit di comunicazione, dipendenza dalle routine, forte sensibilità ai cambiamenti ed alle stimolazioni sensoriali, difficoltà a mantenere il contatto oculare, difficoltà nell'apprendimento, presenza di stereotipie e comportamenti inappropriati. L'estrema variabilità della sindrome rende ogni persona con autismo unica.

Secondo le stime fornite dai Centers for Disease Control and Prevention for Autism, negli Stati Uniti, 1 bambino su 54 è stato diagnosticato con ASD (CDC, 2016). Diversi studi attestano la prevalenza in Europa tra l'1 ed il 2% della popolazione (ASDEU, 2018). Considerando l'incidenza rilevante dell'autismo, è molto importante avviare interventi educativi e comportamentali efficaci fin dai primi anni di vita, al fine di sfruttare la maggiore plasticità del cervello del bambino (Baron-Cohen et al., 2009; Myers et al., 2007).

La scuola risulta quindi uno degli attori principali per la messa a punto di strategie e metodologie personalizzate da applicare in tempi precoci, allo scopo di sfruttare le capacità personali e rispondere alle esigenze di ciascuna persona con piani di apprendimento

modellati sulle sue competenze ed esigenze.

Molti studi attestano l'effetto positivo della musica nel trattare sintomi specifici del disturbo dello spettro autistico (Hallam, 2015). Una revisione della letteratura del 2011 (Simpson e Keen, 2011) evidenzia effetti su comunicazione, socializzazione e comportamento. Per le persone con autismo però può essere difficile imparare il linguaggio complesso della musica e/o suonare uno strumento per la difficoltà che spesso incontrano nel prestare attenzione, decodificare le note e il tempo, nel coordinamento motorio, ecc.

Generalmente i ragazzi, compresi quelli con autismo, hanno una forte attrazione nei confronti della tecnologia che risulta di solito intrinsecamente rinforzante, rassicurante nella sua ripetitività e richiede meno coinvolgimento emotivo rispetto al rapporto interpersonale. Questi aspetti positivi della tecnologia, uniti al diffondersi degli strumenti digitali nella vita quotidiana, hanno incoraggiato numerosi studi volti a supportare l'apprendimento e la riabilitazione mediante l'introduzione di strumenti tecnologici (Golestan et al., 2018; Valencia et al., 2019).

Il progetto Suoniamo contribuisce a questa esperienza innovativa di didattica digitale, con l'obiettivo di favorire il raggiungimento di traguardi educativi nell'area musicale sfruttando le tecnologie digitali. Tale apprendimento dovrebbe anche favorire risultati positivi sia nella vita di relazione e sociale, aumentando in modo collaterale l'autostima, l'integrazione e la socializzazione delle persone con autismo all'interno del contesto scolastico, e sia il raggiungimento di autonomie nella vita quotidiana e, più in generale, un migliore adattamento al contesto.

1. Metodologie di apprendimento di Suoniamo

L'applicazione Suoniamo è stata progettata in collaborazione con professionisti esperti nella formazione e nell'insegnamento musicale di persone con autismo. Proprio considerando il target, essa è stata realizzata per essere accessibile e personalizzabile, sia a livello di interfacce utente sia a livello delle tecniche di insegnamento per l'apprendimento musicale utilizzate, che sono:

- natural learning, per stimolare il cambiamento in modo naturale;
- incidental learning, per fornire occasioni di apprendimento in contesti ludici;
- peer imitation, coinvolgendo i pari nel processo;
- generalizzazione in nuovi contesti, permettendo di connettere l'applicazione ad una tastiera midi per avviare un processo di familiarizzazione con lo strumento reale.

L'apprendimento è strutturato, con un ambiente semplice, ripetibile, prevedibile e coerente che favorisce la costruzione delle competenze e aiuta a familiarizzare con task a difficoltà progressiva contenendo l'ansia.

Le interfacce sono semplici ed intuitive, arricchite da elementi aumentativi che possono essere abilitati o disabilitati a seconda delle esigenze del singolo studente.

2. L'applicazione Suoniamo

Dal punto di vista tecnologico, Suoniamo è una piattaforma ICT, basata su tecnologie Web, che offre esercizi digitali con grado crescente di difficoltà, per imparare a suonare una

tastiera (o pianoforte) e al tempo stesso ad apprendere le basi della teoria musicale, come decodificare le note sul pentagramma, imparare la loro durata (incluse le pause), eseguirle su una tastiera virtuale, apprendere semplici giri musicali.

L'applicazione Suoniamo prevede tre tipi di attività: Moduli Didattici, Suono libero, e Supporto all'esecuzione musicale. Tutte utilizzano una tastiera di pianoforte virtuale arricchita per essere accessibile e personalizzabile, ad esempio aggiungendo il colore e il nome della nota su ciascun tasto.

Le attività e i contenuti educativi possono essere personalizzati dall'insegnante per adeguare le proposte didattiche alle esigenze e preferenze dello studente. Una volta acquisiti gli elementi di base per poter eseguire una melodia, l'insegnante assegnerà allo studente le partiture da imparare. L'obiettivo è quello di facilitare il riconoscimento e l'interpretazione

Fig. 1
Applicazione Suoniamo
Tastiera del pianoforte

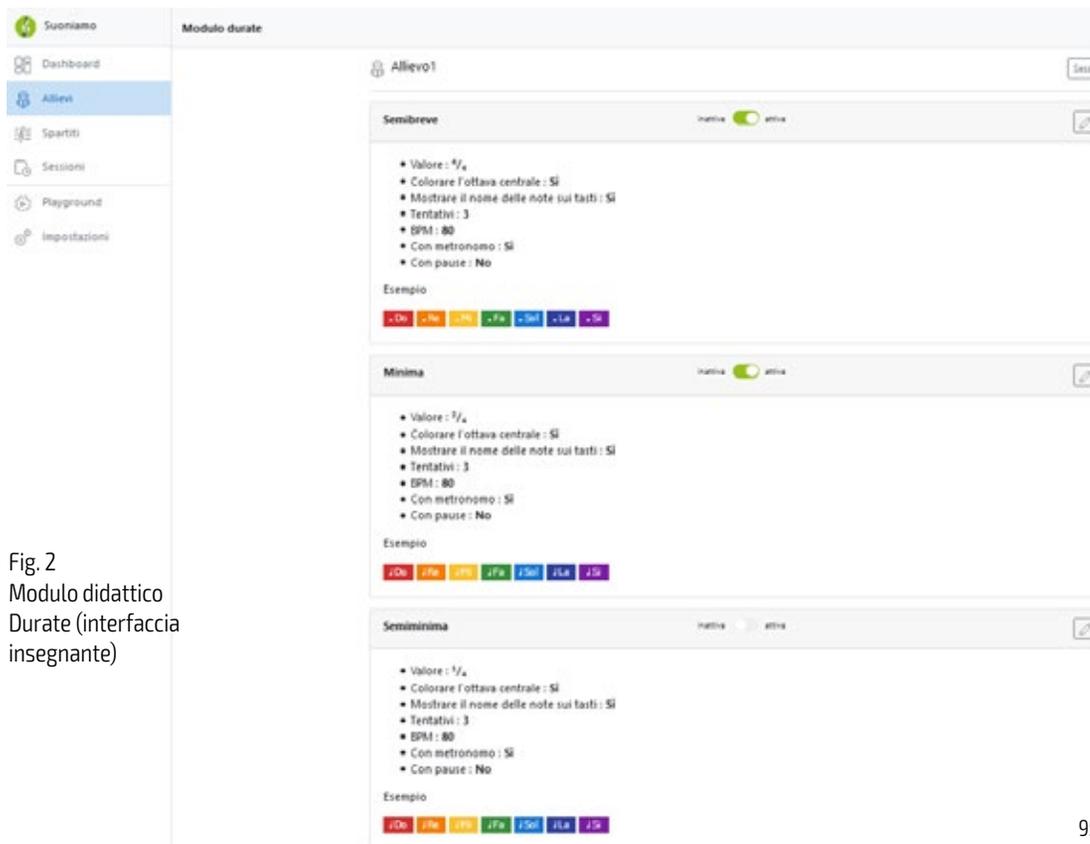


Fig. 2
Modulo didattico
Durate (interfaccia
insegnante)



Fig. 3
Modulo didattico Durate
(interfaccia studente)

della notazione musicale e di guidare via via i ragazzi con autismo alla corretta esecuzione di semplici brani musicali per l'esecuzione collaborativa insieme ai compagni di classe, allo scopo di favorire l'inclusione.

Dopo la fase di apprendimento strutturato, il ragazzo può generalizzare i concetti applicandoli in modo graduale a tastiere fisiche: dapprima una tastiera midi, che può essere collegata via porta USB al tablet su cui gira l'applicazione, e in seguito al vero pianoforte, quando avrà acquisito sufficiente autonomia rispetto al supporto digitale.

3. Sperimentazione e risultati preliminari

La progettazione e lo sviluppo dell'applicazione hanno previsto alcuni cicli di test pilota dei moduli didattici da parte degli utenti finali, coinvolgendo in totale 17 ragazzi con autismo. L'obiettivo era quello di verificare via via l'intuitività e usabilità dell'applicazione ed avere feedback che hanno guidato in modo iterativo la definizione delle interfacce utente e degli elementi aumentativi da utilizzare. Inoltre, i dati osservazionali raccolti durante i test con gli utenti sembrano confermare la validità dell'approccio proposto.

È in programma anche la sperimentazione per un intero anno scolastico dell'applicazione in 4 scuole medie inferiori distribuite sul territorio nazionale: Roma, Lucca, Cellamare (BA) e Mondovì (CN) ma ad oggi, visto che la pandemia da Covid-19 ha costretto le scuole a lunghi periodi di didattica a distanza, e che per gli alunni con autismo è necessario il supporto degli insegnanti di musica e di sostegno nella prima fase di familiarizzazione con l'applicazione, è stato possibile effettuare solo alcune sessioni di utilizzo dell'applicazione Suoniamo e si è rimandata la sperimentazione vera e propria al prossimo anno scolastico 2021-2022, sperabilmente.

In ogni caso, i commenti degli insegnanti e degli studenti, seppure su poche sessioni di test, sono stati generalmente molto positivi, indicando che l'applicazione Suoniamo sembra essere strumento utile per favorire e potenziare l'apprendimento in ambito musicale negli studenti con autismo.

4. Ringraziamenti

Il progetto è stato finanziato da Registro .it (<https://www.nic.it/it>).

Ringraziamo la Dott.ssa Loredana Martusciello (IIT-CNR) per la gestione sistemistica

del server che ospita l'applicazione, e la Dott.ssa Beatrice Rapisarda (IIT-CNR) per la predisposizione delle icone ed il supporto nel progetto grafico dell'applicazione Suoniamo. Ringraziamo inoltre tutti i ragazzi con autismo che ci hanno aiutato nella progettazione partecipativa e durante il test dei diversi prototipi del sistema.

Riferimenti bibliografici

APA - American Psychiatric Association (2013), Diagnostic and Statistical Manual of Mental Disorders, DSM-5.

ASDEU - Autism Spectrum Disorders in the European Union (2018)

<http://asdeu.eu/wp-content/uploads/2016/12/ASDEUExecSummary27September2018.pdf>

Baron-Cohen S., Scott FJ, Allison C., Williams J., Bolton P., Matthews FE, and Brayne C. (2009), Prevalence of autism-spectrum conditions: UK school-based population study, *The British Journal of Psychiatry* 2009; 194: pp 500–509.

CDC - Centers for Disease Control and Prevention (2016), Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016 https://www.cdc.gov/mmwr/volumes/69/ss/ss6904a1.htm?s_cid=ss6904a1_w

Golestan S., Soleiman P., Moradi H. (2018), A Comprehensive Review of Technologies Used for Screening, Assessment, and Rehabilitation of

Autism Spectrum Disorder. arXiv eprint 1807.10986.

Hallam S. (2015), *The power of music: a research synthesis of the impact of actively making music on the intellectual, social and personal development of children and young people*. International Music Education Research Centre.

Myers S.M., Johnson C.P., the Council on Children With Disabilities (2007), Management of Children With Autism Spectrum Disorders. *Pediatrics* (120); pp 1162-1182, DOI: 10.1542/peds.2007-2362.

Simpson K., Keen D. (2011), Music interventions for children with autism: narrative review of the literature. *Journal of autism and developmental disorders*, (41-11), pp 1507-1514.

Valencia K., Rusu C., Quinones D., Jamet E. (2019), The Impact of Technology on People with Autism Spectrum Disorder: A Systematic Literature Review. *Sensors*, (20):4485. doi: 10.3390/s19204485. PMID: 31623200; PMCID: PMC6832622.

Autori



Maria Claudia Buzzi claudia.buzzi@iit.cnr.it

Primo Tecnologo presso l'Istituto di Informatica e Telematica del CNR.

Laureata in Scienze dell'Informazione, si occupa di progettazione partecipativa e realizzazione di servizi ICT innovativi, coordina progetti di ricerca e tecnologici negli ambiti HCI, accessibility, e-learning, e-health.

Docente in corsi universitari e di formazione. Autore di circa cento pubblicazioni internazionali e nazionali. <https://www.iit.cnr.it/claudia.buzzi/>

Marina Buzzi marina.buzzi@iit.cnr.it

Primo Tecnologo presso l'Istituto di Informatica e Telematica del CNR, coordina il gruppo HCT (Human-Centered Technologies) <https://www.iit.cnr.it/marina.buzzi/>. Autore di più di cento pubblicazioni, è responsabile scientifico e/o coordinatore di progetti tecnologici e di ricerca applicata in ambito accessibilità, HCI, eHealth, eLearning, eGovernment, Social Good e Internet degli Oggetti (IoT).

<https://scholar.google.it/citations?user=-z82pfoAAAAJ>



Marco Maugeri maugeri_marco@hotmail.com

È nato a Catania. Si è laureato in lettere alla Sapienza di Roma. Suona il piano, la chitarra, l'ukulele e il banjo, non esclude il giorno in cui finalmente suonerà il violino. Diciamo che su questo contava sui figli. Si occupa di letteratura e storia su quotidiani e televisione fra il 2000 e il 2010. Lavora a scuola da 15 anni, insegna italiano e storia in un liceo artistico.

Gabriella Paolini gabriella.paolini@garr.it

Coordina le attività di Training e E-Learning del GARR. Ha coordinato la partecipazione di GARR nel progetto europeo "UP2University", il progetto di formazione "Progress in training" ed i rapporti con la comunità di scuole collegate alla rete GARR. Si occupa di alfabetizzazione digitale sia in ambito lavorativo che in azioni di volontariato. Laureata in Arti e Scienze dello Spettacolo all'Università La Sapienza di Roma.



Alessandra Sbragia alessandra.sbragia@gmail.com

Ha conseguito la Laurea in Psicologia presso l'Università "La Sapienza" di Roma. Nel corso di più di venti anni di esperienza professionale ha maturato importanti competenze nell'ambito del Disturbo dello Spettro Autistico, utilizzando le principali tecniche di intervento cognitivo-comportamentale e collaborando con numerose associazioni attive sul territorio laziale. Attualmente è socia fondatrice di un'associazione che partecipa ai progetti personalizzati Dopo di Noi.

Caterina Senette caterina.senette@iit.cnr.it

Ha conseguito la Laurea Magistrale in Ing. Biomedica e il Ph.D. in Ing. Informatica presso l'Università di Pisa. Ricercatrice presso CNR-IIT. Coautrice di numerose pubblicazioni in ambito Human-Computer Interaction. Ha esperienza decennale nella estrazione della conoscenza in domini User-Centered, da sempre interessata ai paradigmi di apprendimento umano applicabili ai sistemi informativi. È attualmente coinvolta anche nello studio delle interazioni multiutente tipiche delle reti sociali.



Amaury Trujillo amaury.trujillo@iit.cnr.it

Laurea quadriennale in Ing. Telematica in Messico e laurea magistrale in Ing. dei Sistemi Informativi in Francia. Ha esperienza pluriennale nella ricerca applicata per la progettazione e lo sviluppo di servizi e applicazioni web innovativi, con particolare attenzione a usabilità, accessibilità ed esperienza utente. Attualmente è Ricercatore presso lo IIT-CNR nell'ambito dello Human-Centered Computing e lavora sull'interazione utente tramite l'utilizzo di tecnologie web.

Quando la Scuola cambia, un modello sostenibile

Angelo Canio D'Alessio, Pasquale Boezio

Istituto Comprensivo Statale "Nicola Ronchi" Cellamare (BA)

Abstract. L'esperienza di "didattica a distanza" vissuta nell'anno scolastico 2019/20 ha valorizzato gli interventi di apprendimento significativi, anche attraverso l'affermazione delle tecnologie digitali e delle loro applicazioni in ambito formativo che hanno cambiato la scuola. Si descrivono di seguito le azioni intraprese per erogare il servizio scolastico a distanza a tutti gli studenti dell'Istituto Comprensivo Statale "Nicola Ronchi" di Cellamare modulate attraverso la proposizione di questionari di rilevazione statistica, rivolti a docenti, genitori ed alunni, al fine di comprendere la reale efficacia del sistema formativo digitale promosso e di valutare il raggiungimento degli obiettivi prefissati

Keywords. Didattica, Digitale, Scuola

Il contesto e la piattaforma di erogazione delle attività didattiche

Una scuola aperta è in grado di cogliere le opportunità offerte dalla dimensione internazionale dell'innovazione, instaurando un percorso di cambiamento basato sul dialogo e sul confronto reciproco. Queste opportunità sono scaturite, nel corso dell'anno scolastico 2019/20, a causa dell'emergenza sanitaria COVID-19 che ha imposto scelte di massima flessibilità come risposta ad esigenze che hanno richiesto il superamento di schemi apparentemente intoccabili come la classe, la cattedra, la comunicazione verbale e le verifiche formali.

L'Istituto, così, per rispondere a queste esigenze, ha trasformato gli ambienti fisici in ambienti immateriali, grazie all'uso della piattaforma digitale Google, selezionata perché rispondente alle necessità e capace di offrire funzioni avanzate a supporto della didattica. La gestione della piattaforma ha consentito di sincronizzare e far convergere i prodotti e i risultati di una complessa dinamica di costruzione collaborativa di conoscenze.

Si è fatto maggiormente ricorso ai compiti di realtà e alle esperienze di flipped classroom. Si è limitato lo schema classico della lezione frontale preferendo testi brevi, schemi essenziali, esemplificazioni operative, non escludendo eventuali lavori di ricerca, relazioni su compiti pratici ed esperimenti. La trasformazione dell'aula da spazio fisico chiuso ad ambiente di apprendimento aperto ha modificato il ruolo del docente che è divenuto mentore che accompagna, che dà punti di riferimento, che istruisce sulle metodologie con cui muoversi nell'immaterialità con l'obiettivo di rendere gli studenti sempre più autonomi e responsabili e di offrire loro gli strumenti adatti a un uso consapevole delle risorse per raggiungere i traguardi di competenza.

Il mobile device ha supportato e generato tale cambiamento. In esso sono confluite gran parte delle risorse utili al ragazzo per sviluppare il proprio apprendimento, rappresentando la porta che immette nel cyberspazio.

L'apporto delle famiglie

L'idea di scuola nuova che ne scaturisce non prescinde dall'uso integrato delle tecnologie che diviene sempre più critico man mano che diminuisce il grado di scolarità. L'intermediazione dei genitori diventa spesso indispensabile a causa della scarsa disponibilità di tempo e della necessità di condividere gli stessi strumenti informatici utilizzati dagli adulti per il loro lavoro. Di fronte a una didattica per competenze basata soprattutto sulle nuove tecnologie, si sono promosse attività per lo sviluppo di autonomia e responsabilità degli alunni, di concerto con i genitori, che hanno potuto conoscere a fondo il cambiamento metodologico affrontato, comprendere quando il dispositivo digitale viene usato per scopi personali-ludici e quando per fini didattici, ponendosi come compagni di viaggio dei propri figli.

La partecipazione, secondo l'età e l'organizzazione familiare, la presa in carico degli animali domestici o l'aiuto ai fratelli minori o a vicini che abbiano particolari necessità, sono tutte attività che, accompagnate da un'opportuna riflessione e condivisione, costruiscono competenze diverse e necessarie alla vita personale e sociale e che, nell'ordinario, sono spesso sottovalutate e finanche sottratte all'esperienza dei bambini e dei giovani.

La presenza dei genitori, in particolar modo nei momenti di DaD, è diventata una variabile importantissima dalla quale non è possibile prescindere e che ha imposto nuove forme di corresponsabilità educativa e didattica.

L'Istituto, dunque, si è aperto a nuove forme di collaborazione e comunicazione per incontrare i nuovi bisogni ed esigenze. I genitori sono stati chiamati in causa in un ruolo diverso, di supporto alla continuità della didattica, che li ha impegnati a partecipare, come mai prima, alla complessità dei processi di apprendimento dei loro figli.

I sondaggi

Affinché la scuola funzioni bene e raggiunga gli obiettivi che si pone è necessario anche chiedersi se queste azioni stiano funzionando e come possano migliorarsi per adattarsi a bisogni ed esigenze che non rimangono statiche ma mutano nel tempo. Si è, pertanto, deciso di avviare una consultazione inviando semplici questionari a famiglie, alunni e docenti.

Alle famiglie è stato chiesto se lo studente seguisse con costanza le attività proposte dalla scuola, se fosse autonomo e se fossero in grado di supportare il proprio figlio nelle attività a distanza. Quindi si è chiesto loro un giudizio sulle attività di DaD, sincrone e asincrone, nonché un consiglio per il miglioramento delle attività future.

Il questionario rivolto agli studenti chiedeva se avessero mezzi adeguati per seguire le attività di DaD, se riscontrassero problemi tecnici, se la ritenessero efficace per il proprio percorso formativo, se i docenti avessero dovuto utilizzare la DaD con modalità differenti, chiedendo, infine, un giudizio sulle attività sincrone e asincrone.

Ai docenti, invece, si è chiesto in che modo avessero utilizzato le applicazioni messe a disposizione dalla piattaforma digitale d'Istituto e con quale frequenza, se avessero utilizzato altri sistemi esterni alla piattaforma digitale, se fossero state avviate attività didattiche specifiche per alunni con bisogni educativi speciali e se gli alunni, in generale, avessero avuto problemi nella fruizione delle attività a distanza a causa della scarsa connettività,

quali attività fossero state svolte e quali forme di valutazione adottate.

Dal sondaggio è emerso che gli alunni partecipano con costanza alle lezioni, con una mancata autonomia (supportata dai genitori) che va dal 6% della Secondaria di I grado al 24% della Scuola dell'Infanzia (fig. 1), che le famiglie sono in buona parte soddisfatte delle lezioni sincrone e della loro funzionalità ai percorsi di apprendimento, con una soddisfazione maggiore nella Secondaria di I grado (74%) (fig. 2).

È emerso, inoltre, che gli alunni non hanno particolari problemi nel disporre di mezzi adeguati, che ritengono la Dad efficace al proprio percorso formativo (89%) e più in generale esprimono un giudizio largamente positivo sullo svolgimento delle metodologie di DaD mediante l'uso della piattaforma didattica (99%) (fig. 3).

A conferma di ciò, si riporta che i docenti usano ampiamente i principali strumenti della piattaforma Google, con l'uso di metodologie sincrone e asincrone che, a distanza di 4 settimane dall'avvio della DaD supera ampiamente le tradizionali metodologie (fig. 4).

Fig. 1
Questionario Genitori
Lo/a studente/essa svolge in
maniera autonoma le attività
di DaD proposte dalla scuola
(risposte da 1- per nulla a
5- completamente)

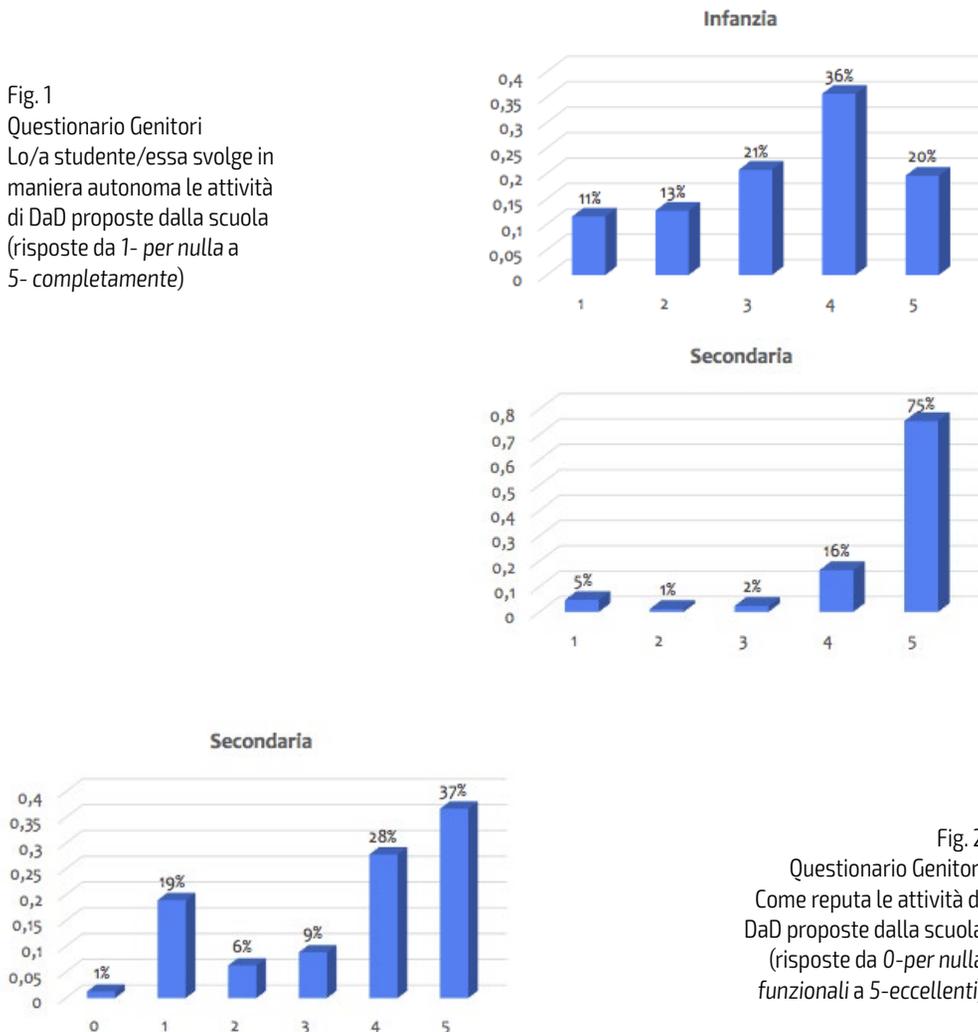


Fig. 2
Questionario Genitori
Come reputa le attività di
DaD proposte dalla scuola
(risposte da 0-per nulla
funzionali a 5-eccellenti)

Fig. 3
Questionario studenti
La didattica a distanza mi sembra efficace per il mio percorso formativo (a sx) e Giudizio sullo svolgimento delle metodologie di DaD mediante l'utilizzo di piattaforme didattiche (a dx)

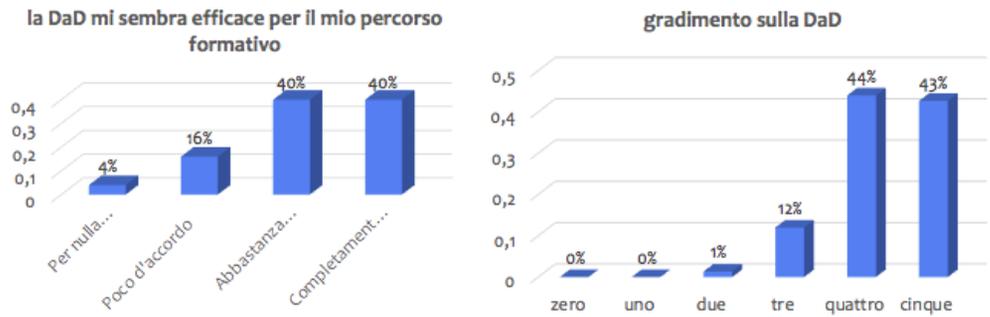
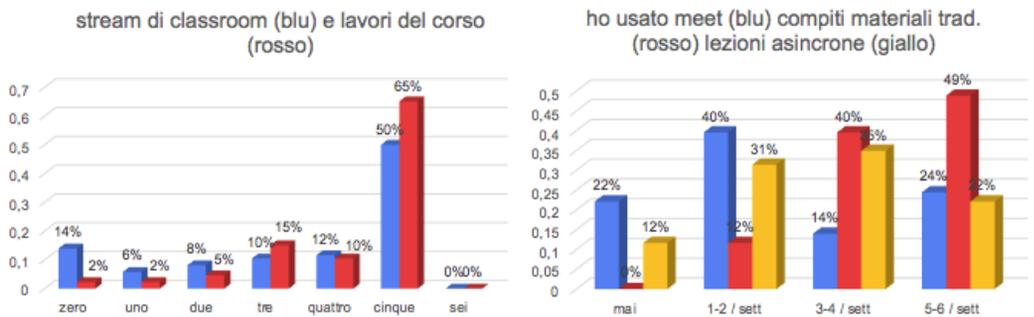


Fig. 4
Questionario docenti
Uso settimanale degli strumenti della piattaforma LMS.



Conclusioni

Tutto ciò porta a considerare positivo l'intervento della scuola sia in termini di infrastruttura che di organizzazione del lavoro; infatti, lì dove si utilizzano correttamente gli strumenti più innovativi consentiti dalle tecnologie di DaD le risposte, sia delle famiglie sia degli studenti, si attestano chiaramente su livelli più che soddisfacenti (figg. 2 e 3) L'attività di DaD, non sostitutiva di quella in presenza, si intersecherà stabilmente in futuro con quest'ultima provocando un cambiamento di paradigma che disegnerà la Scuola del Futuro. Sarà una Scuola più attenta alle relazioni, sarà una Scuola nella quale si riconoscerà ai docenti un ruolo, che forse si era perso per strada, di educatori e mentori e non di contenitori da cui estrarre tutto lo scibile umano. Con la Didattica a Distanza tutta la società si sta rendendo conto, attraverso il riconoscimento dell'atto intenzionale da parte dei docenti di avere cura dei propri figli, della centralità della Scuola nel mondo contemporaneo.

Autori

Angelo Canio D'Alessio acdlessio@libero.it

Nato a Bari nel '73. Laurea in Fisica nel '98, Dottorato di Ricerca in Fisica nel 2002. Pubblica i risultati degli studi su riviste e atti di convegni. Insegna Fisica nella Scuola Superiore dal 2001 al

2019 rivestendo il ruolo di referente d'Istituto per GARR ed EDUROAM. Dal 2019 diviene Dirigente Scolastico dell'Istituto Comprensivo "Nicola Ronchi" di Cellamare (BA).



Pasquale Boezio pasquale.boezio@istitutoronchi.edu.it

Si laurea in Architettura nel 2008. Specializzato in "Architettura digitale", dal 2010 al 2015 si occupa di progettazione architettonica, con particolare attenzione ai temi della rappresentazione grafica, della modellazione digitale e della fotografia. Dal 2016 è docente di "Arte e Immagine" presso la Scuola Secondaria di I grado. Esperto in "Tecnologie della didattica".

Didattica Online: modello di formazione per docenti di ogni ordine e grado

Cecilia Fissore, Francesco Floris, Marina Marchisio, Matteo Sacchet,
Sergio Rabellino

Università degli Studi di Torino

Abstract. Nei quadri di riferimento delle istituzioni scolastiche di ogni ordine e grado viene sottolineata l'importanza di far sviluppare agli studenti le competenze digitali e di promuovere l'uso delle tecnologie a scopo educativo. Ciò richiede ai docenti un insieme di competenze digitali sempre più ampio e sofisticato per far fronte a richieste in rapida evoluzione. Questa necessità diventa ancora più impellente in periodi di emergenza come quello per la pandemia da COVID-19, in cui diventa urgente e indispensabile predisporre una didattica online con gli studenti. Operando nel contesto del framework DigCompEdu, l'Università di Torino ha sviluppato un modello per la formazione dei docenti, dalla scuola primaria all'Università, sulla didattica online. In questo lavoro viene presentato il modello in relazione al framework insieme ad alcuni risultati ottenuti

Keywords. Competenze digitali, Comunità di apprendimento, Didattica online, Formazione docenti, Tecnologie digitali

Introduzione

Nei quadri di riferimento delle istituzioni scolastiche di ogni ordine e grado viene sottolineata l'importanza di far sviluppare agli studenti le competenze digitali e la promozione delle tecnologie a scopo educativo. La competenza digitale è definita come la capacità di saper utilizzare con dimestichezza e spirito critico le tecnologie. Ogni cittadino necessita di competenze digitali per poter beneficiare delle opportunità digitali date dalla digitalizzazione della nostra società, ma anche per mitigarne i possibili rischi. Il dovere di aiutare gli studenti a diventare digitalmente competenti richiede ai docenti di sviluppare un insieme di competenze digitali sempre più ampio e sofisticato, che diventa obbligo in periodi di emergenza come quello per la pandemia da COVID-19. La chiusura delle scuole e delle Università, e la conseguente sospensione delle attività didattiche in presenza, ha reso fondamentale trovare alternative per garantire continuità didattica agli studenti. L'attualità di questa emergenza enfatizza la necessità di un'adeguata formazione del personale docente che opera a tutti i livelli d'istruzione.

In ambito internazionale e nazionale sono stati sviluppati numerosi quadri di riferimento, strumenti di autovalutazione e programmi di formazione per descrivere gli aspetti della competenza digitale per i docenti e aiutarli a valutare le loro competenze digitali, identificare e rispondere alle loro esigenze di formazione. DigCompEdu (Redecker, 2017) è rivolto a docenti a tutti i livelli di istruzione e mira a fornire un quadro di riferimento gene-

rale utile per coloro che si occupano di modelli capaci di far sviluppare competenze digitali, descrivendo come le tecnologie digitali possono essere usate per innovare l'insegnamento e l'apprendimento. DigCompEdu, nella formulazione attuale, propone sei aree (Fig. 1) che spaziano dal coinvolgimento professionale alla competenza digitale degli studenti; ogni area è articolata in diverse competenze per un totale di 22.

Nel contesto di questo quadro teorico, l'Università di Torino ha analizzato il proprio modello per lo sviluppo di competenze digitali per la Didattica Online, tramite la formazione degli insegnanti di ogni ordine e grado. Tale modello può essere proposto per la formazione di tutti i docenti, dalla scuola primaria all'università.

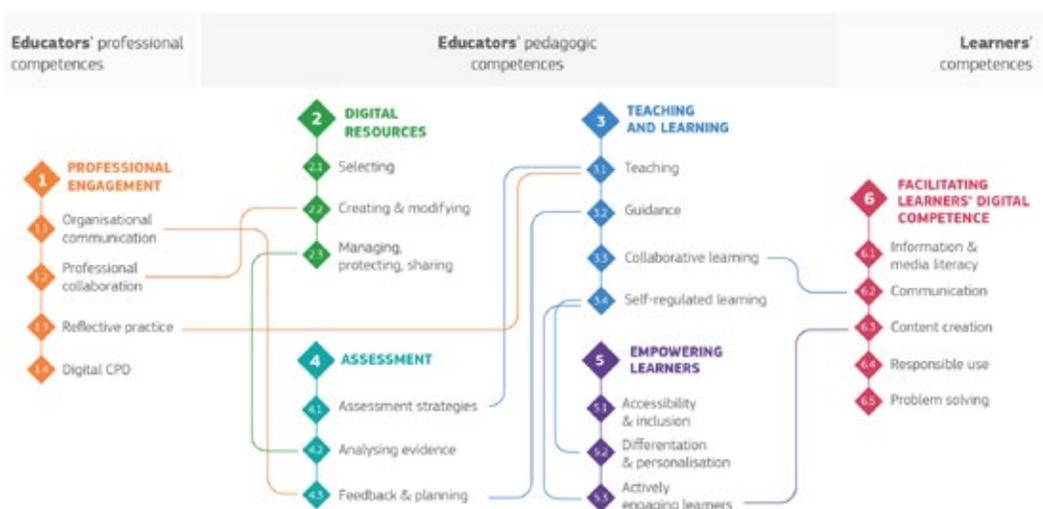


Fig. 1
DigCompEdu
framework,
aree e connessioni

1. Il modello

Una componente importante per la Didattica Online è l'Ambiente Digitale di Apprendimento, un ecosistema di apprendimento in cui insegnare, apprendere e sviluppare competenze in classe, online in modalità mista o ibrida. L'ambiente è composto da una componente umana, una componente tecnologica e dalle relazioni tra le due. È importante sottolineare che con il termine Didattica Online non si intende soltanto le attività formative sincrone fra gli utenti (videoconferenza, chat, etc.), ma anche l'apprendimento asincrono (che non necessita la presenza di docente e studente nello stesso momento): forum di discussione, consegna compiti, condivisione di materiali, workshop, test con valutazione automatica, accesso ai materiali didattici. Il modello dell'Università di Torino è basato sull'utilizzo di una piattaforma Moodle, integrata con un servizio di web-conference, un sistema di valutazione automatica e un Ambiente di Calcolo Evoluto (ACE) particolarmente utile per lo studio delle discipline STEM. Esso coinvolge, come evidenziato di seguito, tutte le 6 aree tematiche del DigCompEdu.

Il modello propone una formazione ai docenti che li aiuti a sfruttare le nuove potenzialità tecnologiche, a guidare gli studenti nell'utilizzo delle tecnologie e ad aggiornarsi in modo continuo (sottoarea 1.4) (Brancaccio et al., 2019). Nel formare i docenti

all'utilizzo di nuove tecnologie è fondamentale farli collaborare fra loro, creando una comunità di apprendimento, per scambiare conoscenze ed esperienze (sottoarea 1.2). I docenti possono così condividere risorse digitali (sottoarea 2.3) e imparare a identificare quelle adatte agli obiettivi di apprendimento, al gruppo di studenti e allo stile di insegnamento (sottoarea 2.1). La formazione dei docenti riguarda le metodologie didattiche e la progettazione di attività con l'uso delle tecnologie digitali, utili nelle diverse fasi del processo di apprendimento.

Una delle metodologie proposte è quella del problem solving (sottoarea 6.5), con situazioni contestualizzate in cui gli studenti devono applicare conoscenze e competenze per risolvere problemi. I docenti imparano a personalizzare l'ambiente digitale di apprendimento per i propri studenti e organizzare contenuti online. L'utilizzo della piattaforma integrata e di risorse digitali supporta i processi di insegnamento centrati sullo studente e l'apprendimento collaborativo (sottoarea 3.3) e il coinvolgimento attivo degli studenti (sottoarea 5.3). Attraverso la piattaforma integrata è possibile enfatizzare l'aspetto interattivo delle risorse digitali. I docenti imparano a creare risorse digitali e, quelli di discipline scientifiche, ad usare un ACE per esplorare una situazione e generalizzare la risoluzione di un problema (sottoarea 2.2). L'integrazione della piattaforma con un sistema di valutazione automatica basato su un ambiente di calcolo, consente la creazione di domande con parametri variabili e a risposta aperta, che permettono di valutare il processo e non solo il risultato (sottoarea 4.1).

In particolare, l'Università di Torino ha sviluppato e sperimentato con successo un modello per la valutazione formativa automatica (Barana et al., 2018) con feedback interattivi e immediati, capace di promuovere l'insegnamento adattivo e l'apprendimento autoregolato (sottoarea 4.3). Queste strategie permettono una didattica personalizzata capace di adattarsi alle necessità di ogni singolo studente (sottoarea 5.2). Il docente infine può monitorare costantemente i comportamenti degli studenti in piattaforma e interpretare tutti i dati sull'apprendimento di ogni singolo discente (sottoarea 4.2).

2. Conclusioni

Il modello per la Didattica Online è stato proposto per la formazione di docenti di ogni ordine e grado. All'interno del progetto nazionale PP&S (Fissore et al., 2020) (www.progettopp.it), promosso dal Ministero dell'Istruzione, docenti delle scuole secondarie di primo e secondo grado di tutta Italia collaborano insieme a distanza e ricevono una formazione online continua. All'inizio dell'emergenza COVID-19, nel mese di marzo 2020, è stata organizzata una formazione sincrona online straordinaria di due incontri settimanali, in particolare per i nuovi docenti iscritti al progetto, alla quale hanno partecipato quasi 400 docenti. Attualmente sulla piattaforma sono presenti oltre 28000 utenti, di cui circa 2000 docenti. Gli incontri di formazione sono proseguiti lungo tutto il 2020 e il 2021 con cadenza settimanale. Il modello è stato utilizzato anche per la formazione di docenti universitari e di docenti appartenenti alla scuola primaria della regione Piemonte, nell'ottica di promuovere un curriculum verticale.

Riferimenti bibliografici

Barana, Conte, Fioravera, Marchisio, Rabellino, 2018, A Model of Formative Automatic Assessment and Interactive Feedback for STEM, IEEE 42nd Annual Computer Software and Applications Conference, pp. 1016-1025.

Brancaccio, Esposito, Marchisio, Pardini, Sacchet, 2019, Open professional development of math teachers through an online course, Proceedings of the International Conference on e-Learning in MCCSIS 2019, pp. 131-138.

Fissore, Marchisio, Rabellino, 2020, Secondary School Teacher Support and Training for Online Teaching during the COVID-19 Pandemic, European Distance and E-Learning Network (EDEN) Proceedings, pp. 311-320.

Redecker, C. 2017, European framework for the digital competence of educators: DigCompEdu, Publications Office of the European Union.

Autori



Cecilia Fissore cecilia.fissore@unito.it

Cecilia Fissore, laureata in Matematica, è dottoranda in Digital Humanities presso le Università di Torino e Genova. Il progetto di ricerca riguarda lo sviluppo di metodologie digitali per l'apprendimento delle lingue, caratterizzate dall'utilizzo di un ambiente digitale di apprendimento integrato con un sistema di valutazione formativa automatica. Collabora in numerosi progetti di ricerca nel campo della Digital Education. È membro del DELTA Research Group - Digital Education for Learning and Teaching Advances - gruppo di ricerca dell'Università degli Studi di Torino.

Francesco Floris francesco.floris@unito.it

Francesco Floris lavora attualmente presso il Dipartimento di Biotecnologie Molecolari e Scienze della Salute dell'Università degli Studi di Torino. I principali interessi di Francesco sono lo studio dei Learning Analytics e la ricerca di soluzioni data-driven per migliorare l'esperienza degli studenti in un Ambiente Digitale di Apprendimento. È membro del DELTA Research Group - Digital Education for Learning and Teaching Advances - gruppo di ricerca dell'Università degli Studi di Torino.



Marina Marchisio marina.marchisio@unito.it

Marina Marchisio è professoressa ordinaria di Matematiche Complementari e Delegata del Rettore per lo sviluppo e la promozione della Digital Education di Ateneo. La sua ricerca è focalizzata sull'apprendimento e l'insegnamento delle discipline matematiche e STEM con metodologie digitali e tecnologie innovative. Coordina il Delta Research Group - Digital Education for Learning and Teaching Advances - gruppo di ricerca dell'Università degli Studi di Torino. È membro del gruppo di lavoro Problem Posing & Solving del Ministero dell'Istruzione. Coordina vari progetti di ricerca, di didattica, di e-learning, per l'orientamento universitario, a favore del successo formativo e contro la povertà educativa. È la responsabile scientifica di Orientamente e Start@unito. È autrice di numerose pubblicazioni nell'ambito della Digital Education.



Sergio Rabellino sergio.rabellino@unito.it

Sergio Rabellino è tecnico di ricerca presso il Dipartimento di Informatica. Collabora con i gruppi di ricerca in Sicurezza, Informatica, Elaborazione ad alte prestazioni, Intelligenza artificiale ed E-learning. È sviluppatore Moodle e architetto hardware/software di piattaforme di e-learning, cloud e HPC. È responsabile tecnico delle piattaforme moodle Start@Unito, Orient@amente, iLearn e PPS e ha scritto oltre 40 pubblicazioni su strumenti e metodi di e-learning. È membro del DELTA Research Group - Digital Education for Learning and Teaching Advances - gruppo di ricerca dell'Università degli Studi di Torino.

Matteo Sacchet matteo.sacchet@unito.it

Matteo Sacchet è assegnista di ricerca presso il Dipartimento di Biotecnologie Molecolari e Scienze per la Salute. I suoi interessi di ricerca riguardano l'insegnamento e l'apprendimento della Matematica e, in generale, delle discipline STEM utilizzando le tecnologie digitali e metodologie innovative, integrate in un ambiente di apprendimento digitale. Si occupa di progetti di e-learning e corsi online di carattere open. È membro del DELTA Research Group - Digital Education for Learning and Teaching Advances - gruppo di ricerca dell'Università degli Studi di Torino.



Una piattaforma scalabile per esami sicuri da remoto

Damiano Perri ^{1,2}, Osvaldo Gervasi ², Marco Simonetti ^{1,2}

¹Università di Firenze, ²Università di Perugia

Abstract. Durante il periodo di pandemia una delle maggiori criticità che ha coinvolto, seppur per ragioni diverse studenti e docenti, è stata la corretta valutazione delle conoscenze. In molti casi la prima reazione a seguito delle chiusure imposte per ragioni sanitarie è stata quella di imporre esami in modalità esclusivamente orale. Presentiamo quindi un applicativo web chiamato LibreEOL che permette lo svolgimento da remoto di esami di livello universitario in modalità sicura. L'applicativo è in grado anche di monitorare e prevenire comportamenti scorretti da parte dello studente, nel rispetto della normativa europea vigente in materia di trattamento dei dati sensibili (GDPR)

Keywords. e-assessment, open source, web application, GDPR, cloud computing

Introduzione

Affrontare una sessione di esame orale da remoto può nella maggior parte dei casi risultare un'attività particolarmente stressante e complessa, soprattutto se il numero degli studenti è elevato.

Il nostro gruppo di ricerca ha sviluppato nel corso degli ultimi venti anni, sugli esiti di un progetto finanziato dalla Comunità Europea nell'ambito dell'iniziativa Leonardo (denominato DASP - Distance Assessment System for accreditation of competencies and skills acquired through in company Placements, 1997-2000), un sistema di valutazione elettronica per la verifica delle conoscenze e competenze acquisite in seguito a stage formativi. Nel corso degli anni il software è stato continuamente potenziato(1). Nel 2015 la European Chemistry Thematic Network (ECTN) ha adottato tale piattaforma per la realizzazione degli esami EchemTest dedicati alla valutazione delle competenze chimiche nei vari ambiti della disciplina.

Con l'avvento della pandemia da COVID-19 ci siamo focalizzati sull'implementazione di una tecnologia efficace specifica per realizzazione remota di esami in modalità sicura e volta a monitorare e prevenire comportamenti scorretti da parte degli studenti, sempre nel rispetto della normativa europea vigente in materia di trattamento dei dati sensibili (GDPR). A tal fine il sistema acquisisce una serie di dati sensibili (foto e suoni), i quali vengono conservati il tempo necessario affinché il docente possa validare la prova di esame e successivamente vengono rimossi. I dati sono disaccoppiati dall'identità degli studenti ai quali fanno riferimento e sono salvati in modo codificato fuori dallo spazio di lavoro web. Il prodotto è stato reso disponibile ai docenti dell'Università di Perugia a partire da novembre 2020, in forma di prototipo; già da fine dicembre, dopo che il Rettore aveva emanato un decreto nel quale autorizzava l'uso di LibreEOL per l'esecuzione di esami scritti da

remoto, si registrava una notevole attività di utilizzo della piattaforma.

Dalla fase iniziale alla data del 02 Settembre 2021 si sono registrati nella piattaforma 385 docenti, sono state create 601 materie di esame, 24109 domande, 2772 sessioni di esame e sono stati sostenuti 35361 esami dagli studenti, con un larghissimo consenso da parte sia dei docenti che degli studenti (sentiti attraverso i rappresentanti nel Consiglio degli Studenti e tramite singoli studenti che ci hanno contattato, oltre agli esiti del questionario compilato dagli studenti e da quello compilato dai docenti). Il sistema può essere aperto alle Università interessate.

1. Struttura dell'applicazione

LibreEOL è una Web App, strutturata con un backend basato su PHP, un frontEnd basato su JavaScript e come RDBMS viene utilizzato MySQL. Il codice sorgente è disponibile in GitHub alla url <https://github.com/DamianoP/LibreEOL>.

Nello sviluppo del codice abbiamo sempre cercato di adottare soluzioni standard e prediletto approcci nativi HTML5, oltre a rispettare le buone pratiche in ambito Human Computer Interaction e User eXperience. L'esecuzione di un test da parte di uno studente non comporta scaricamento di componenti, a eccezione dei cookies tecnici. A seguito dell'adozione della piattaforma da parte del nostro Ateneo, l'applicazione, basata su container Docker, è stata migrata su un ambiente cloud commerciale; questo ha permesso di realizzare la scalabilità orizzontale e una serie di ottimizzazioni proprietarie per massimizzare le prestazioni.

Durante lo svolgimento dell'esame lo studente viene monitorato attraverso i seguenti controlli:

- Monitoraggio dell'audio ambientale: grazie a questa tecnologia, analizziamo i picchi sonori catturati dal microfono del PC dello studente e registriamo l'audio ambientale, fino al decremento dell'intensità sonora al di sotto di una soglia definita. Questo consente di rilevare eventuali voci che suggeriscono.
- Monitoraggio del mouse e della finestra attiva del browser: la finestra di esame viene forzata a schermo intero e ogni tentativo di uscita o di click al di fuori della stessa viene segnalato e conteggiato.
- Monitoraggio del volto: tramite una rete neurale, estraiamo i landmarks relativi al volto dello studente; grazie ad essi, siamo in grado di rilevare la direzione del suo viso e riconoscere i momenti in cui si sta "distraendo" dall'esame. Tale evento viene conteggiato e contemporaneamente viene anche scattata una foto allo studente.
- Timelapse: con cadenza periodica di 4 secondi, viene scattata una foto allo studente; questa tecnica ci permette di ricostruire la dinamica temporale dell'intera sessione di esame.
- Live View: il docente può monitorare contemporaneamente tutti gli studenti partecipanti, attraverso una pagina web; inoltre, da essa, il docente può interagire con gli studenti, inviandogli messaggi broadcast o personali. Un esempio di pagina del Live View è mostrato in Figura n.1

Durante e dopo l'esame, il docente potrà accedere alla lista degli iscritti all'esame (che nel caso di un esame completato indicherà anche il voto suggerito dal sistema), come mostra-

to in Figura n.2, dalla quale potrà accedere alla pagina “Live View”, monitorare l’andamento delle distrazioni, controllare le informazioni raccolte dal sistema attraverso una pagina riportante i dati del candidato, la foto iniziale e quella del documento, il Timelapse, la lista delle distrazioni video e audio. Da qui il docente potrà decidere se validare l’esame o meno.

Fig. 1
Pagina del live view, attraverso la quale il docente può vedere e interagire in tempo reale con gli studenti

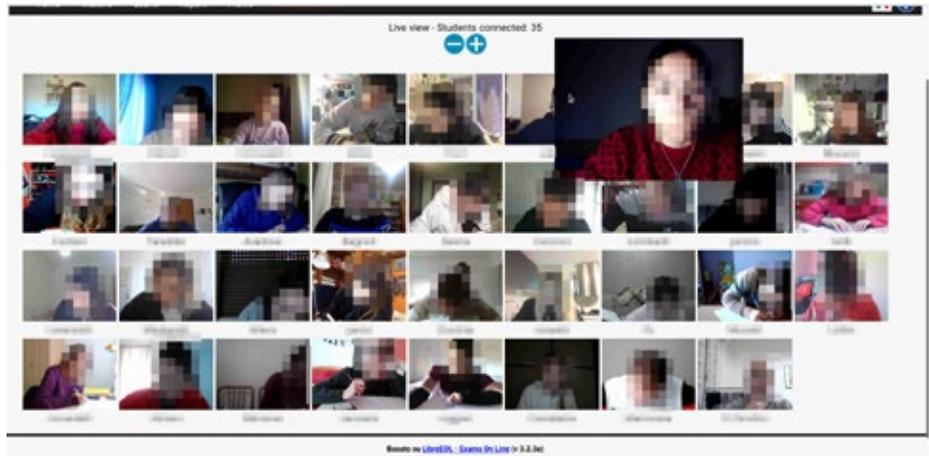


Fig. 2
Lista iscritti all’esame e altre informazioni correlate, utili per il docente

Visualizzati 110 studenti registrati

Stato	Cognome *	Nome	E-mail	09-06-2021 09:53	09-06-2021 10:42:07	Tempo totale 04:42:04	Punteggio (max 30) 18,9	Punteggio finale (max 30) 17,7	Eventi Video 0	Eventi Audio 0	Eventi Mouse 0	Pagina scorribile 0	Controllo
	2021-07-27 10:01:17	2021-07-27 11:01:47	00:59:30	18	18	0	0	0	0	
	2021-07-27 10:01:17	2021-07-27 11:00:12	00:59:55	21,7	22	0	0	0	0	
	2021-07-27 10:01:17	2021-07-27 10:59:45	00:58:28	25,7	26	0	11	0	0	
	2021-07-27 10:59:53	2021-07-27 10:59:57	00:00:04	26,6	26	0	0	0	0	
	2021-07-27 10:59:12	2021-07-27 10:59:16	00:00:04	27,6	28	0	0	0	0	
	2021-07-27 10:01:02	2021-07-27 11:01:06	01:00:04	27	27	0	3	0	0	
	2021-07-27 10:01:12	2021-07-27 11:01:16	01:00:04	28,9	24	0	74	0	0	
	2021-07-27 10:52:30	2021-07-27 10:52:34	00:00:04	29	29	0	0	0	0	
	2021-07-27 10:00:40	2021-07-27 11:00:47	01:00:07	29,1	29	0	2	0	0	

Test consegnati: 106

Chiedi Assistenza Richiedi Guida Live view Reporti Appoggia

2. Organizzazione del materiale del docente

Il docente per ciascuna materia predisporrà argomenti e domande, classificate come facili, medie, difficili, è possibile dare un peso differente alle tre tipologie. Le domande possono essere di tipo “risposta singola” (solo una risposta vera), “risposta multipla” (più risposte vere che concorrono al punteggio complessivo), “Sì/No”, “Vero/Falso”, “Risposta aperta” (è atteso un input che verrà giudicato dal docente), “Numerica” (atteso un risultato numerico), “Testo” (si sceglie una o poche parole), “Hotspot” (lo studente deve selezionare un punto in un’immagine), “QR Code” (lo studente scrive su uno o più fogli l’elaborato, che verrà poi caricato in modo sicuro attraverso il cellulare e l’uso di un QR Code).

Le domande e le risposte possono essere inserite in diverse lingue. In base alle impostazioni dello studente le domande di esame verranno presentate nella lingua dello studente o nella lingua di riferimento se non disponibile la lingua dello studente.

3. Conclusioni

Il sistema LibreEOL risulta uno strumento facile da usare sia da studenti che da docenti e consente di introdurre delle buone pratiche nella fase della valutazione degli studenti. Inoltre, permette in modo sicuro e non invasivo di affrontare lo svolgimento di esami da remoto (al di là della pandemia, si pensi a studenti ospedalizzati o costretti a casa, studenti lavoratori, etc, che potrebbero così sostenere esami da remoto in modo sicuro per tutti). Consente di raggiungere gli obiettivi Horizon 2021-2027 in merito alla digitalizzazione della formazione e della didattica. Molteplici sono i vantaggi della digitalizzazione delle prove di esame: il monitoraggio degli studenti, lo studio statistico dell'andamento dei voti, per intercettare eventuali problematiche sommerse anche nell'attività di docenza, la dematerializzazione delle prove, l'affidabilità globale del sistema.

L'uso degli indicatori statistici permette inoltre di migliorare la qualità delle domande e delle risposte.

Riferimenti bibliografici

[1] O. Gervasi and A. Laganà, EoL a web-based distance assessment system. Lecture Notes on Computer Science . 3044, 854-862 (2004)

Autori

Damiano Perri damiano.perri@unifi.it

Svolge il dottorato di ricerca in Informatica nel programma congiunto tra i Dipartimenti di Matematica e Informatica dell'Università di Firenze e dell'Università di Perugia. La sua attività di ricerca è incentrata su GPGPU Computing, ambienti paralleli e distribuiti, realtà virtuale e aumentata, intelligenza artificiale, reti neurali, modelli predittivi per ottimizzare gli operatori più importanti nelle Reti Neurali Convolute (CNN). Dal 2016 è il responsabile tecnico del progetto LibreEOL



Oswaldo Gervasi osvaldo.gervasi@unipg.it

Ph.D., Professore Associato presso il Dipartimento di Matematica e Informatica, Università degli Studi di Perugia. Insegna Architettura Reti, Sistemi di Realtà Virtuale, High Performance Computing e Human Computer Interaction presso il dipartimento di Informatica dell'Università degli Studi di Perugia. Si occupa di Parallel e Distributed Computing, Cloud e High Performance Computing, GPGPU Computing, Intelligenza Artificiale, Realtà Virtuale e Aumentata, e-Learning.



Marco Simonetti m.simonetti@unifi.it

Marco Simonetti si è laureato in Ingegneria Elettronica presso l'Università di Bologna. Ha lavorato nel settore dell'Ingegneria Clinica e ha partecipato con un team di lavoro alla realizzazione di sistemi automatizzati per le Amministrazioni Sanitarie. Lavora come in-



segnante di matematica nelle scuole superiori ed è studente di dottorato in Informatica presso l'Università di Firenze. Tra i suoi interessi ci sono Deep Learning e Reti Neurali, Quantum Computing e Didattica della Matematica.

Reti di apprendenti, reti di apprendimento. Un uso ecologico della social network analysis per la progettazione di LMOOC

Alessandro Puglisi

Università per Stranieri di Siena

Abstract I Language MOOC (LMOOC) sono una realtà in ascesa all'interno del mondo dei MOOC. La diffusione di questa modalità di erogazione della formazione porta con sé enormi opportunità, tanto per lo studioso quanto per il progettista della formazione, ma pone anche sfide di grande rilievo. Il contributo prende le mosse da una ricerca svolta avvalendosi dell'analisi delle reti sociali per indagare le interazioni comunicative in un ambiente di apprendimento online dedicato all'italiano L2. Più nello specifico, si propone una riflessione sullo sviluppo di un framework per la raccolta, il trattamento e l'utilizzo dei dati relativi all'apprendimento online delle lingue in contesti di massa. In tal modo si intende perseguire l'obiettivo di implementare un ciclo di sviluppo della formazione linguistica online fondato sull'analisi dei dati in termini quantitativi, qualitativi e interazionali

Keywords. LMOOC, apprendimento, social network analysis, progettazione

Introduzione

Un report apparso di recente sull'aggregatore Class Central (Shah, 2020b) ha definito il 2020 come “il secondo anno dei MOOC”, dopo il 2012 che aveva visto una vera e propria esplosione di questa modalità di erogazione di contenuti formativi (Pappano, 2012), online, aperta e di massa. Il 2020, a leggere i dati relativi ai MOOC provider, può ben essere definito come uno snodo fondamentale. La crisi sanitaria legata alla diffusione pandemica di SARS-CoV-2 e le conseguenti misure di contenimento del contagio, fra le quali il lockdown, hanno impresso un'enorme, e imprevedibile, spinta alla formazione online. Questa accelerazione si è espressa sia sul fronte di un aumento esponenziale degli apprendenti, che per quanto riguarda gli investimenti delle piattaforme (Shah 2020a) e una loro sempre maggior capacità di monetizzazione. In più, se guardiamo, ancora sulla base del report di Class Central, agli ambiti più richiesti, in termini di MOOC, dopo lo scoppio della pandemia, possiamo rilevare come alcuni settori trasversali, attinenti alle cosiddette soft skills, per esempio personal development e management & leadership, abbiano guadagnato terreno su altri ambiti caratterizzati invece in maniera più specifica, e legati a precise capacità professionali. D'altro canto, nella top ten delle materie più quotate tra gli utenti dei MOOC, è da rilevare l'ingresso delle communication skills e quello delle lingue straniere.

1. I Language MOOC

La nostra proposta si muove proprio sul terreno dell'insegnamento e apprendimento del-

le lingue straniere in contesti pensati per grandi e grandissimi numeri di apprendenti. I LMOOC (Language MOOC) hanno ovviamente una storia più giovane di quella dei MOOC tout court, costituendone peraltro una sottocategoria. È evidente come il crescente interesse, da parte dei potenziali apprendenti, per lo studio delle lingue in Rete attraverso i MOOC, stia sollecitando risposte concrete, tanto in termini di analisi scientifiche dell'apprendimento linguistico in contesti di massa, quanto per ciò che riguarda la progettazione di corsi online di questo tipo. Far fronte a queste esigenze significa perciò confrontarsi con enormi quantità di dati, trattabili con le metodologie e gli strumenti della data science, che consentono di integrare indagini quantitative, qualitative e analisi delle interazioni.

2. Una proposta ecologica

Quella che proponiamo qui è una riflessione, nel metodo e nel merito, sulle possibilità di raccolta, trattamento e utilizzo dei dati nell'ambito della formazione linguistica in Rete in maniera ecologica. Intendiamo qui l'aggettivo nella sua accezione più propria, guardando quindi ai nostri contesti di insegnamento e apprendimento come si guarda a degli ecosistemi.

La nostra riflessione parte da una ricerca condotta sul LMOOC Introduction to Italian dell'Università per Stranieri di Siena, corso online di italiano della durata di sei settimane, ospitato dal provider FutureLearn. Il progetto è stato svolto tra l'Università per Stranieri di Siena e la University of Toronto, dal 2017 al 2020 nell'ambito del dottorato di ricerca in Linguistica e didattica dell'italiano a stranieri dell'Università per Stranieri di Siena, grazie a una borsa di studio Pegaso vincolata, finanziata dalla Regione Toscana per studi legati al mondo dell'e-learning.

In particolare, è stata condotta un'analisi delle interazioni comunicative intraprese dagli utenti di Introduction to Italian (Run 1) attraverso la funzionalità di commento presente all'interno delle singole pagine del LMOOC. Sfruttando il software R e il package {igraph}, a partire dagli oltre centomila commenti inseriti dagli utenti, sono stati prodotti dei grafi, orientati e pesati, delle relazioni comunicative instauratesi. A tale scopo si è guardato a tre endpoint, posti dopo la prima settimana di corso, dopo la terza settimana e a fine corso. Sulle strutture dati costruite si sono poi calcolate alcune misure di centralità, fra le quali betweenness centrality e closeness centrality, oltre a indici come authority score e hub score. La ricerca ha consentito, anzitutto, di dare avvio a una narrazione diversa dei dati relativi all'insegnamento e apprendimento linguistico online in contesti di massa; una narrazione in cui lo studente si avvia a riguadagnare davvero il centro della scena. In più, il lavoro di ricerca ha permesso di evidenziare prospettive di analisi dei dati di apprendimento tramite strumenti di machine learning per individuare predittori positivi e negativi dell'outcome degli apprendenti.

3. Il framework

Considerando quanto detto sopra, si sono poste le basi, dunque, per un framework che preveda la raccolta dei dati a partire da sistemi di tracciamento xAPI allo scopo di costruire matrici di adiacenza. A partire da esse, si potranno produrre grafi in cui i singoli nodi sono

rappresentati non solo da utenti, ma anche da documenti, pagine Web, attività interattive, in linea con la luminosa intuizione connettivista (Siemens, 2004). Il framework viene proposto in relazione all'erogazione di corsi online di lingua italiana a stranieri; tuttavia, è evidente come il modello sia applicabile ad ambiti formativi differenti, a patto che i relativi corsi siano svolti in Rete.

In tal senso, i punti-chiave del framework che proponiamo si possono così sinteticamente rappresentare in dieci fasi reiterate ciclicamente, in un'ottica SAM (Successive Approximations Model):

1. erogazione di un corso di italiano L2 tramite piattaforma aperta (CMS) con sistemi di tracciamento xAPI;
2. raccolta dei dati di apprendimento (learning analytics) in formato interoperabile xAPI;
3. esportazione dei dati in formato *.csv (comma-separated values);
4. importazione dei *.csv nel software di analisi statistica R;
5. costruzione di grafi orientati e pesati tramite {igraph};
6. analisi dei grafi costruiti in termini di centralità statistica in relazione agli outcome degli apprendenti;
7. applicazione di tecniche di machine learning per l'individuazione di valori predittivi rispetto all'outcome;
8. identificazione di possibili migliorie a livello progettuale;
9. implementazione delle migliorie all'interno del corso;
10. nuova erogazione (con ritorno alla fase 1).

4. Conclusioni

Attraverso il framework esposto, crediamo di poter fornire un contributo per rispondere alle nuove e pressanti sfide poste al mondo della formazione linguistica online nella presente congiuntura storica. Approcciare tale fenomeno, che emerge dalla Rete, con un atteggiamento scientifico che sia davvero "di Rete", sembra costituire la strada migliore per superare la tendenza a produrre indagini aneddotiche, ponendo le basi per un utilizzo efficace ed efficiente di nuove cornici di lavoro.

Riferimenti bibliografici

Pappano L. (2012, novembre 2), The Year of the MOOC, URL: <https://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html> [11/09/2021].

Shah D., (2020a, 2 maggio), How Different MOOC Providers are Responding to the Pandemic, URL: <https://www.classcentral.com/report/mooc-providers-response-to-the-pandemic/> [11/09/2021].

Shah D., (2020b, 14 dicembre), The Second Year of the MOOC: A Review of MOOC Stats and Trends in 2020, URL:

<https://www.classcentral.com/report/the-second-year-of-the-mooc/> [11/09/2021].

Siemens G., (2004, 12 dicembre), Connectivism: A Learning Theory for the Digital Age,

URL: https://www.academia.edu/2857237/Connectivism_a_learning_theory_for_the_digital_age [11/09/2021].

Autori



Alessandro Puglisi puglisi@unistrasi.it

Alessandro Puglisi ha ottenuto il Dottorato di ricerca in Linguistica e didattica dell'italiano a stranieri nel 2021 presso l'Università per Stranieri di Siena, dove è attualmente borsista di ricerca. È stato International Visiting Graduate Student presso la University of Toronto (Department of Italian Studies).

Fra i suoi interessi di ricerca: insegnamento e apprendimento della lingua italiana online, analisi delle reti sociali di apprendimento, uso delle tecnologie digitali in glottodidattica.

Riconessioni: competenze digitali e nuove infrastrutture per abilitare l'innovazione dal basso del paradigma educativo e realizzare pienamente l'autonomia scolastica

Giovanni Luca Spoto, Marcello Enea Newman, Veronica Ruberti, Chiara Ciociola
Fondazione per la Scuola della Compagnia di San Paolo

Abstract. Presentazione delle caratteristiche chiave di Riconessioni, modello di innovazione sociale della Compagnia di San Paolo, realizzato a Torino dalla Fondazione per la Scuola, volta a mettere a fuoco quali elementi infrastrutturali e quali percorsi di rafforzamento delle competenze digitali dei docenti possano aiutare le scuole del primo ciclo a fare un uso più consapevole e deciso della propria autonomia

Keywords. Didattica digitale, Strategie e politiche di sviluppo per infrastrutture digitali, Formazione e trasferimento tecnologico, Innovazione didattica, Riconessioni

Introduzione

L'autonomia che caratterizza le scuole e i docenti nel sistema scolastico italiano è un elemento potenzialmente abilitante del cambiamento: soprattutto per quello che riguarda le metodologie di insegnamento, i docenti possono fare virtualmente qualsiasi cosa ritengono opportuna quando sono nella propria aula. Le scuole hanno autonomia sull'uso e sulla suddivisione del tempo, (INDIRE, a2021) sull'uso degli spazi (INDIRE, b2021), e addirittura sulla scelta dei contenuti disciplinari da affrontare. Uno dei principi dietro l'autonomia è che questa possa permettere ai docenti di cogliere agilmente le opportunità offerte dall'innovazione metodologica e tecnologica dell'istruzione; più di quanto sarebbe possibile in un sistema più centralizzato. L'autonomia dovrebbe anche aiutare le scuole a rispondere alle emergenze, come le chiusure rese necessarie dal COVID-19 nel 2020 e nel 2021.

Benché il sistema italiano contenga al suo interno casi eccezionali, ampiamente discussi e celebrati, in cui singole scuole si sono distinte per un approccio olistico e trasversale all'innovazione, diversi fattori indicano che il sistema nel suo insieme non stia riuscendo né a fare i cambiamenti di lungo termine necessari per fornire agli studenti di oggi le competenze di cui avranno bisogno nel mondo di domani, né ad attuare cambiamenti di breve termine necessari per affrontare emergenze come il COVID-19.

Secondo l'indice DESI, solo il 22% degli italiani possiedono competenze digitali avanzate (European Commission, 2021) (paragonato alla media UE del 33%). Questo dato va letto in due modi: in primis ci dice che il sistema educativo italiano non è riuscito a fare gli aggiustamenti necessari per fornire ai cittadini le competenze di cui avranno bisogno in

futuro; inoltre, ci dice che è improbabile che i docenti stessi posseggano le competenze digitali necessarie per trarre il massimo dall'innovazione, né che siano in grado di sviluppare queste competenze nei propri alunni.

Il progetto Riconessioni, realizzato a Torino dalla Fondazione per la Scuola della Compagnia di San Paolo, ha speso la triennalità dal 2017 al 2020 conducendo un'iniziativa cittadina sull'innovazione educativa bottom-up abilitata dalle TIC che ha ingaggiato più di 1500 docenti in laboratori intensivi di formazione professionale e connesso più di 246 plessi scolastici alla fibra ottica. Riconessioni rappresenta un modello per sviluppare l'autonomia e il protagonismo dei docenti attorno alla pedagogia digitale: nel lungo termine, per abilitare un cambio dal basso del paradigma educativo, nel breve termine, per accrescere la resilienza del sistema. I risultati del progetto a Torino dimostrano come l'innovazione metodologica della formazione docenti e un'infrastruttura di rete adeguata possano permettere alle scuole di innovare anche radicalmente le proprie pratiche didattiche. Seguono le caratteristiche chiave dell'infrastruttura di rete, con un affondo sulle azioni emergenziali attuate durante la pandemia COVID-19 per aiutare le scuole a garantire la continuità didattica.

1. Fibra ottica e accesso a Internet

Il modello di rete è organizzato per livelli, di cui la rete metropolitana in fibra ottica si colloca primo livello di priorità. A Torino, Open Fiber ha realizzato un'infrastruttura metropolitana dedicata alle scuole coinvolte, interamente basata su fibra ottica, organizzata secondo una topologia a stella.

TOP-IX, Torino Piemonte Internet Exchange, ha avuto un ruolo strategico, veicolando il traffico IP da e verso le scuole con collegamenti simmetrici diretti a 10Gbps, così come altri punti di interscambio, e garantendo l'interconnessione tra gli operatori per ridurre la latenza.

Le scuole coinvolte in Riconessioni sono state veicolate sulla Rete Italiana dell'Università e della Ricerca Scientifica, denominata comunemente "Rete GARR". Le caratteristiche tecniche della rete GARR permettono di avere limitati ritardi di trasmissione: questo la rende particolarmente adatta non solo a trasmettere grandi quantità di dati, ma anche alla comunicazione in tempo reale e alla trasmissione audio/video ad altissima definizione (Cisco, 2017).

2. Le scuole coinvolte a Torino e nei comuni dell'hinterland

Lo sviluppo dell'infrastruttura di rete si concluderà nel secondo quarto del 2021 e al suo termine conterà 246 scuole. Sono state connessi tutte i plessi raggiungibili dalla rete a Torino e nei grandi comuni dell'hinterland. La rete in fibra ottica di Riconessioni raggiunge oggi oltre l'80% degli studenti di Torino e hinterland, oltre il 90% nella città di Torino.

3. Osservazioni sul traffico della rete

Le chiusure rese necessarie dal COVID-19 nel 2020 e 2021 hanno portato gli studenti e docenti a spostare le lezioni online. Dall'analisi del traffico Internet in transito sull'infra-

struttura di rete di Riconessioni dall'inizio dell'anno scolastico 2020/2021 a marzo, si è riscontrato un notevole aumento del traffico sia in download sia in upload.

Si è tenuto conto, per ciascun giorno, dei valori medi e dei valori massimi, riscontrando una forte crescita dalla seconda settimana di marzo e registrando incrementi del traffico fino al 250% rispetto all'andamento standard, con mantenimento del traffico aumentato nelle settimane successive.

Tali valori rendono evidente come sia fondamentale e indispensabile dotare le scuole di infrastrutture di rete solide e ad alte prestazioni, almeno per garantire la continuità didattica in momenti emergenziali.

4. Reti di scuole

Le scuole che hanno partecipato al progetto Riconessioni nella città di Torino si sono riunite in Reti per migliorare il sistema formativo, potenziare e integrare il servizio scolastico e per ottimizzare le risorse, in particolare attorno alla connettività. Le Reti Riconessioni sono attualmente quattro e raggruppano a oggi circa sessanta scuole collocate sul territorio della Città Metropolitana di Torino. La nascita delle Reti Riconessioni mostra come la partecipazione comune alla rete in fibra ottica possa fare da innesco per collaborazioni di altro tipo.

5. La formazione

Dopo una prima triennalità di progetto in cui la formazione si è concentrata soprattutto su percorsi laboratoriali tenuti in presenza che hanno coinvolto più di 1.553 docenti e Dirigenti Scolastici (e altri 5.000 docenti attraverso la formazione a cascata), Riconessioni ha allargato la sua offerta formativa per poter raggiungere un numero ancora maggiore di docenti attraverso iniziative online e in modalità blended, come descritto nei prossimi paragrafi.

6. Webinar

Dal 9 marzo al 2 luglio 2020 Fondazione per la Scuola, tramite Riconessioni, ha proposto un ricco programma di 44 webinar formativi. 68 ore di diretta rivolte a Dirigenti Scolastici, docenti ed educatori per proporre risorse, strumenti e buone pratiche di Didattica a Distanza (DAD) e non solo. All'inizio dell'anno scolastico 2020-2021 l'offerta di webinar è stata mantenuta, diventando sempre più una piattaforma e un percorso di autoformazione strutturato in registrazioni video, lesson plan da sperimentare, linee guida per progettare, materiali di approfondimento. A oggi, marzo 2021, più di 50.000 persone hanno partecipato ai webinar.

7. Galleria

Durante l'emergenza COVID-19 Riconessioni ha attivato la propria comunità di docenti e organizzato un processo di raccolta e condivisione di attività didattiche innovative, per condividere buone pratiche inclusive e creative ed essere di reciproco stimolo. A oggi la Galleria di Riconessioni conta oltre 150 attività didattiche con materiali, video e sitogra-

fia di riferimento che possono essere scaricate e sperimentate in un nuovo contesto scolastico. A livello nazionale, la Galleria è la più estesa raccolta di attività didattiche prodotte dai docenti.

8. Pensiero Computazione e Apprendimento Creativo

“Pensiero computazionale e apprendimento creativo” è un percorso formativo per docenti delle scuole del primo ciclo distribuito su un arco temporale di 3 mesi. Alterna momenti sincroni di sperimentazione collaborativa a momenti asincroni di approfondimento e attività individuali. Il cuore del percorso è l'apprendimento della metodologia dell'Apprendimento Creativo, sviluppato dal Lifelong Kindergarten del MIT Media Lab. Questo modello di formazione unisce digitale e analogico, presenza e distanza, sincrono e asincrono, a una forte spinta alla collaborazione e al confronto tra docenti, ponendo le basi per un uso più consapevole, radicale e sostenibile dell'autonomia di cui i docenti dispongono.

Riferimenti bibliografici

Cisco. (2017), Supporting a Digital Transformation in Schools, eSchoolNews.com

European Commission. “Italy in the Digital Economy and Society Index” digital-strategy.ec.europa.eu. <https://ec.europa.eu/digital-single-market/en/scoreboard/italy> (ultimo accesso 23 agosto 2021)

INDIRE. “Uso flessibile del tempo (compattazione)” [indire.it](http://innovazione.indire.it/avanguardieeducative/uso-flessibile-tempo). <http://innovazione.indire.it/avanguardieeducative/uso-flessibile-tempo> (ultimo accesso 23 agosto 2021) – a

INDIRE. “Aule laboratorio disciplinari” [indire.it](http://innovazione.indire.it/avanguardieeducative/aule-laboratorio-disciplinari). <http://innovazione.indire.it/avanguardieeducative/aule-laboratorio-disciplinari> (ultimo accesso 23 agosto 2021) – b

Autori



Giovanni Luca Spoto giovanni.spoto@fondazioneeducazione.it

Ingegnere informatico e ricercatore dal 2009 nel settore ICT. Esperto in reti e tecnologie per Riconessioni, un progetto della Compagnia di San Paolo realizzato dalla Fondazione per la Scuola.

Marcello Enea Newman marcello.newman@fondazioneeducazione.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.



Veronica Ruberti veronica.ruberti@fondazioneeducazione.it

Veronica si occupa di didattica innovativa, formazione docenti, comunicazione e community engagement per il progetto Riconessioni. Ha conseguito una laurea in Scienze Cognitive all'Università di Trento e un master in Comunicazione della Scienza alla SISSA di Trieste. Ha raccontato curiosità scientifiche ai microfoni di Radio Capodistria e scrive di divulgazione scientifica per l'infanzia.





Chiara Ciociola chiara.ciociola@fondazione scuola.it

Coordina la formazione docenti per Riconessioni. Ha lavorato come consulente al MI per l'attuazione del Piano Nazionale Scuola Digitale. Ha coordinato A Scuola di OpenCoesione, presso la Presidenza del Consiglio dei Ministri, programma nazionale di cittadinanza digitale in didattica ibrida. Scrive di uso critico delle tecnologie su Neural Magazine, rivista di arte e cultura digitale.

