

CONFERENCE
GARR
2023

Saperi interconnessi

Infrastrutture per una scienza
ad alte prestazioni

FIRENZE, 14-16 GIUGNO 2023

SELECTED
PAPERS

 Consortium
GARR

Conferenza GARR 2023 - Saperi interconnessi - Infrastrutture per una scienza ad alte prestazioni
Firenze, 14-16 giugno 2023

ISBN 978-88-946629-2-4

DOI 10.26314/GARR-Conf23-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 2023

Numero di copie: 400

Stampa: Tipografia Graffietti Stampati snc

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

Tutti i materiali relativi alla Conferenza GARR 2023 sono disponibili all'indirizzo:

www.conf23.garr.it

Indice

- 8 **Digital Twin Components for Geophysical Extreme Phenomena: the example of Volcanic Hazards within the DT-GEO project**
Stefano Cacciaguerra, Antonio Costa, Francesca Quareni, Paolo Papale¹, Flavio Cannavò, Arnau Folch, Giovanni Macedonio, Sara Barsotti
- 17 **La piattaforma IEMAP (Italian Energy Materials Acceleration Platform): Il ruolo dei dati aperti e condivisi**
Claudio Ronchetti, Sergio Ferlito, Marco Puccini, Simone Giusepponi, Francesco Buonocore, Massimo Celino, Sara Marchio, Giovanni Ponti
- 24 **The RESTART project: REsearch and innovation on future Telecommunications systems and networks, to make Italy more "smART"**
Cristina E. Costa, Antonella Bogoni, Ilenia Tinnirello, Guido Maier, Giacomo Verticale, Raffaele Bolla
- 30 **The IT@CHA Virtual Lab: a set of enabling technologies for Digital Twin implementation and enrichment in Cultural Heritage**
Marialuisa Mongelli, Beatrice Calosso, Samuele Pierattini, Marco Puccini, Giovanni Ponti
- 38 **IIT Dataverse: il repository dell'Istituto Italiano di Tecnologia per la conservazione e la condivisione dei dati FAIR della ricerca**
Valentina Pasquale, Alessandro Bruchi, Ugo Moschini, Elisa Molinari, Daniele Rossetto, Francesca Cagnoni, Stefano Bencetti
- 43 **FOSSR-Fostering Open Science in Social Science Research: Building Italy's Innovative Open Cloud Infrastructure**
Serena Fabrizio, Mario Sicurezza, Mario Ciampi, Emanuela Reale
- 51 **A Cloud Computing infrastructure for the Competence Centre for the Conservation of Cultural Heritage**
A. Alkhansa, A. Bombini, A. Costantini, L. dell'Agnello, F. Taccetti
- 55 **Toward Effective Cybersecurity Introduction for High-school girls**
Gabriele Costa, Serenella Valiani

- 61 **Progetto IS-LeGI per la didattica digitale e la formazione**
Fabrizio Turchi, Francesco Romano
- 64 **Gemelli digitali e gestione dei dati del patrimonio culturale**
Achille Felicetti, Franco Niccolucci
- 69 **Play INAF: risorse digitali per la didattica dell'astronomia**
Maura Sandri, Rino Bandiera, Silvia Casu, Laura Daricello, Federico Di Giacomo, Mariachiara Falco, Maria Teresa Fulco, Silvia Galletti, Giuliana Giobbi, Laura Leonardi, Marco Malaspina, Claudia Mignone, Barbara Olmi, Serena Pastore, Agatino Rifatto, Rachele Toniolo, Ginevra Trinchieri
- 76 **Il Sistema di Gestione della Sicurezza (SGSL) dell'INAF e integrazione con il database dell'anagrafica dei dipendenti**
Vito Conforti, Francesca Romana Porta, Elisabetta Bartone Giuseppina Antico, Angelo Boccaccini, Sabrina Ciprietti, Gianluigi Deiana, Nicola Di Cicco, Stefano Giovannini, Davide Inamo, Jacopo Lenzi, Giovanni Liggio, Sergio Mariotti, Francesco Massaro, Barbara Neri, Giovanni Occhipinti, Mario Pepe, Francesco Perrotta, Pier Raffaele Platania, Irene Zago
- 84 **Phishing simulato: le campagne di phishing sono tutte equivalenti?**
Fabrizio Fioravanti, Marius Bogdan Spinu, Marco Alamanni
- 88 **Cyber Resilience Act and Open Source: problemi aperti e prospettive future**
Nadina Foggetti
- 93 **BullyBuster - A framework for bullying and cyberbullying action detection by computer vision and artificial intelligence methods and algorithms**
Wanda Nocerino, Vincenzo Gattulli, Stefano Marrone, Giulia Orrù, Grazia Terrone, Antonio Galli, Michela Gravina, Donatella Curtotti, Donato Impedovo, Gian Luca Marcialis, Carlo Sansone
- 99 **Decidere al ritmo degli eventi**
Massimo Carboni
- 104 **I progetti ICSC, il Centro Nazionale HPC, Big Data e Quantum Computing e TeRABIT**
Claudio Grandi, Mauro Campanella



SAPERI INTERCONNESSI

*Infrastrutture per una
scienza ad alte prestazioni*

Comitato di programma

Claudio Allocchio GARR
Claudia Battista GARR
Elis Bertazzon GARR
Debora Berti UNIVERSITÀ DI FIRENZE
Alessandro Brunengo INFN
Simonetta Buttò ICCU
Stefano Cacciaguerra INGV
Mauro Campanella GARR
Massimo Carboni GARR
Raffaele Conte CNR
Andrea Dell'Amico CNR
Giuseppe De Marco ESPERTO DI IDENTITÀ DIGITALI
Sara Di Giorgio GARR
Monica Forni UNIVERSITÀ DI BOLOGNA
Sabato Mellone IRCCS RIZZOLI E UNIVERSITÀ DI BOLOGNA
Gabriella Paolini GARR
Claudio Pisa GARR
Federica Tanlongo GARR
Sabrina Tomassini GARR
Davide Vagheti GARR
Simona Venuti GARR
Carlo Volpe GARR

Tutte le presentazioni e maggiori informazioni
sono disponibili sul sito dell'evento:
www.conf23.garr.it

Digital Twin Components for Geophysical Extreme Phenomena: the example of Volcanic Hazards within the DT-GEO project

Stefano Cacciaguerra¹, Antonio Costa¹, Francesca Quareni¹, Paolo Papale¹, Flavio Cannavò¹, Arnau Folch², Giovanni Macedonio¹, Sara Barsotti³

¹Istituto Nazionale di Geofisica e Vulcanologia (INGV), ²Geociencias Barcelona (GEO3BCN-CSIC), ³Icelandic Meteorological Office (IMO)

Abstract. The project Digital Twin for GEOphysical extremes-(DT-GEO) aims to use Digital Twin Components to create replicas of physical systems, serving as a virtual laboratory to study natural extreme events. The rationale is the intrinsic risks of potentially catastrophic events to anthropic activities, infrastructures, and cultural heritage. In the framework of the project, this paper describes how the DTC workflow architecture is designed, focusing on flexibility, scalability, and maintainability, and how it is further developed. To demonstrate how ICT efforts can expand horizons in Geosciences, an application to volcanic hazard is presented taking as a case study the 2019 volcanic eruption of Raikoke (Kuril Islands).

Keywords. Digital Twin, Geoscience Data Management, Geoscience and ICT, Volcanoes and HPC

1. Introduction

Natural extreme events, like earthquakes, landslides, tsunamis, and volcanic eruptions, may involve significant risks to infrastructures, cultural heritage, and human lives, and, since ever, the scientific community is called to contribute better understand the complexity of their non-linear behavior and to mitigate their impact. On the other hand, today, major efforts are spent in Europe towards the empowering of ICT, innovation and skills development, through initiatives like European Open Science Cloud (EOSC), European Digital Innovation (EDI), and European High Performance Computing (EuroHPC - Ejarque et al., 2022).

Meanwhile, large amounts of multidisciplinary Earth system data are now available in near-real-time according to the FAIR principles (Wilkinson et al., 2016). As we enter the Exascale Era of computation, Geosciences require modeling codes able to perform over massive HPC clusters (Folch et al., 2023).

Recognizing the potential of Digital Twins Components (DTC, e.g. Fuller et al. 2020), the project Digital Twin for GEOphysical extremes-DT-GEO (Carbonell et al. 2023) has been funded under the Horizon Europe programme (2022-2025) within the European Commission initiative Destination Earth in order to develop numerical clones that accura-

tely replicate the extreme geophysical phenomena. The project aims to integrate advancements from European projects, such as EOSC-synergy and eFlows4HPC (Talia et al. 2023), as well as the Center of Excellence for Exascale in Solid Earth-ChEESE (Folch et al. 2023) (Fig.1). The goal is to produce self-contained and containerized DTCs, incorporating cutting-edge codes, artificial intelligence layers, large data streams, and data assimilation methodologies. Serving as virtual replicas, the DTCs will provide a better understanding and more comprehensive analysis of extreme geophysical phenomena.

12 DTCs and their workflows are designed to be optimized on EuroHPC systems or cloud computing environments in order to address volcanic eruptions, tsunamis, earthquakes, and induced seismicity. When driven by near-real-time monitoring data, DTCs can contribute to early warning systems, estimation of potential consequences, urgent computations, and rapid post-event analysis. Alternatively, when driven by user-defined synthetic data, DTCs can draw various scenarios improving hazard assessment.

To validate their functionality, DTCs use selected Site Demonstrators (SD) with data from the EPOS Integrated Centralized Services of the European Plate Observing System (EPOS). The long-term vision is to integrate the DTCs into the Destination Earth framework, facilitating internal component coupling and promoting external integration with other digital twins. Furthermore, DTCs/SDs will be the basis for the EPOS Distributed Integrated Centralized Services (Cocco et al. 2022).

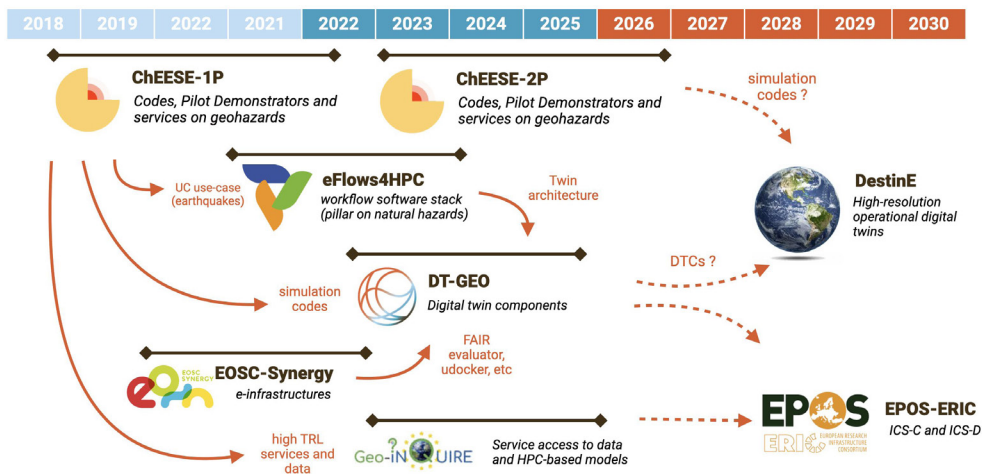


Fig. 1 Background and destination of DT-GEO in the European infrastructural research ecosystem

2. Data Architecture, Workflows and e-Infrastructures

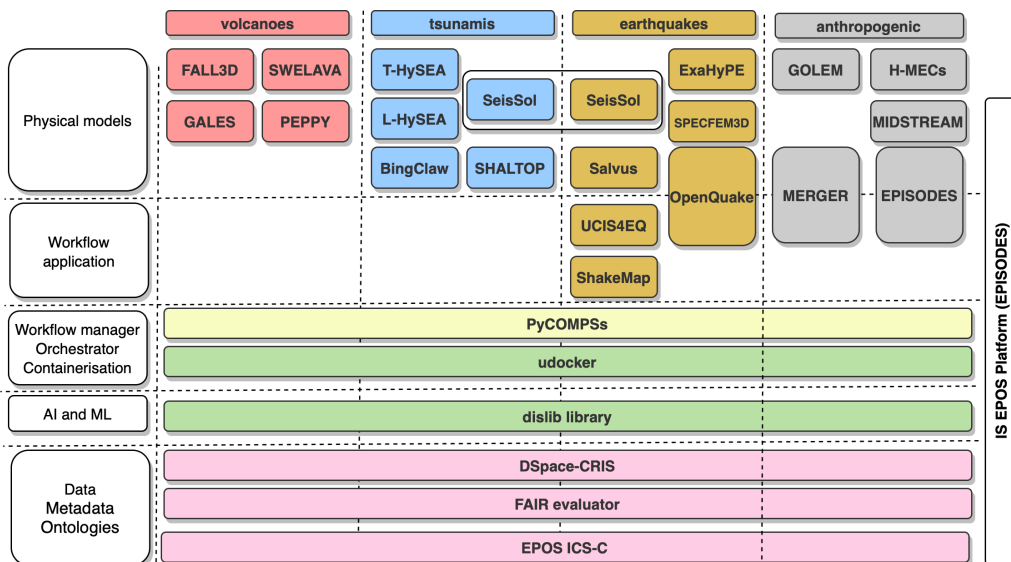
The DTC workflow architecture should focus on flexibility, scalability, and maintainability. To identify the requirements of different DTCs/SDs, it is essential to consider the data ecosystem, such as data and metadata quality, collection methods, processing activities, and storage strategies. The required modular architecture should be adaptable to the execution across various e-Infrastructures, including HPC systems and cloud environments, within federated Research Infrastructures (RI). The architecture must also be easily exten-

dable and deployable.

To develop a DTC, DT-GEO adopts a layered software stack as shown in Fig. 2. On the top there is the physical model describing the specific geophysical phenomenon. Intermediate layers follow, managing workflows, containers, and ensuring data and metadata quality. The physical model is the main component of the workflow, which can in turn enhance it with artificial intelligence, machine learning or data assimilation techniques. This framework accesses a robust ecosystem with quality repositories for software, data, and metadata. Finally, the workflow is containerized to be executed on different e-Infrastructures. To implement the workflows, the PyCOMPSs environment is combined with the software stack developed by the eFlows4HPC project. PyCOMPSs (Tejedor, 2017) is a Python-based framework to develop and execute parallel applications on distributed infrastructures, whereas eFlows4HPC software provides a range of libraries and frameworks for data processing, Machine Learning (ML), visualization, and workflow management. Thanks to PyCOMPSs and eFlows4HPC, developers can create scalable workflows that handle large volumes of data and are easily deployed and maintained.

To enable DT-GEO workflows on e-Infrastructures, such as FENIX (Alam et al. 2022) or EuroHPC, is required to achieve seamless access, to deploy DTCs/SDs as containers and to execute on-demand DTCs/SDs on a HPC or cloud environment. In order to implement this approach, DTCs/SDs are containerized applications downloadable from Dockerhub or other docker registries. DTCs/SDs are implemented with udocker (Gomes et al. 2018), because the containerized images can be downloaded and executed without needing root access, ensuring a secure and streamlined deployment process.

Fig. 2
DT-GEO layered
software stack



3. Application to volcanic phenomena: a case study

Volcanoes encompass many different domains, from deep regions of magma storage below and within volcanoes, to dykes and conduits, surrounding rocks, the Earth surface, and the atmosphere, with non-linear coupling. From each domain a stream of high-quality data (accessible e.g. through the EPOS RI) is produced, which is precious to base research, and feeds risk analyses and civil protection activities, including early warning systems and emergency planning.

The first goal within DT-GEO is the definition of the requirements, specifications, and interoperability of the volcano related components, including software engineering tools, HPC resources, and data lakes. Once the first step is achieved, four DTCs will be developed and implemented for volcanic unrest, volcanic ash clouds and tephra ground accumulation, lava flows, and volcanic gas dispersal.

Here we present as a case study the diffusion in the atmosphere of a volcanic ash column. Fig. 3 shows an aerial photograph of an ash column, and a sketch of the characteristics of the phenomenon, highlighting the observed data.

Modeling volcanic ash concentrations relies on the measurement of various physical parameters that are affected by errors. To enhance the accuracy of assessments in an operational context, the project utilizes geostationary satellite observations (including atmospheric radiance, VIS/IR, and UV/IR with shadow), aerial visual observations, and ground-based measurements like radar, lidar, and VIS/IR cameras. The integration of the observed data into numerical models is able to significantly improve the accuracy of assessments.

As sketched in Fig. 4, the DTC replicates the phenomenon of volcanic ash dispersion in atmosphere by feeding the existing input observations (from ECMWF, ESA, INGV/EPOS, and IMO) to a data assimilation tool based on an ensemble approach, which combines the FALL3D dispersion physical model (Folch et al. 2009) with the Parallel Data Assimilation Framework (PDAF). The FALL3D+PDAF system is designed to operate in parallel and support online-coupled data assimilation, and can be effectively integrated into operational workflows thanks to HPC capabilities.

The spatial distribution of column mass, according to Mingari et al. (2022), is shown in Fig. 5a, 18 hours after the start of the eruption, according to the considered real state, after which synthetic observations are created (Fig.5b). Figs. 5c and 5d show the distribution of column mass estimated from the physical numerical model, and the result of the DTC analysis with data assimilation.

In both cases, the ash cloud is transported eastward by upper-level winds. However, the results obtained without assimilation exhibits a broader spatial distribution compared to the real state due to ensemble dispersion. Furthermore, the modeling without assimilation fails to reproduce the peak mass position of the column occurring in the northern region of the cloud.

In contrast, the analyzed mass load field closely approaches the real state after a few assimilation cycles (Fig.5d).

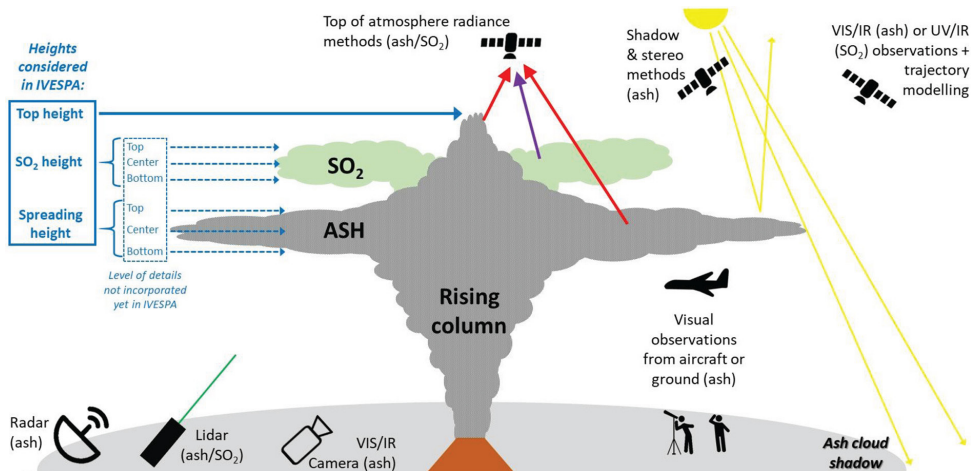


Fig. 3
 Top: Aerial photograph of an ash column generated by a volcanic eruption (Raikoke, Kuril Islands) after Global Volcanism Project (2019). Bottom: sketch highlighting the observed data (bottom) after Aubry et al. (2021)

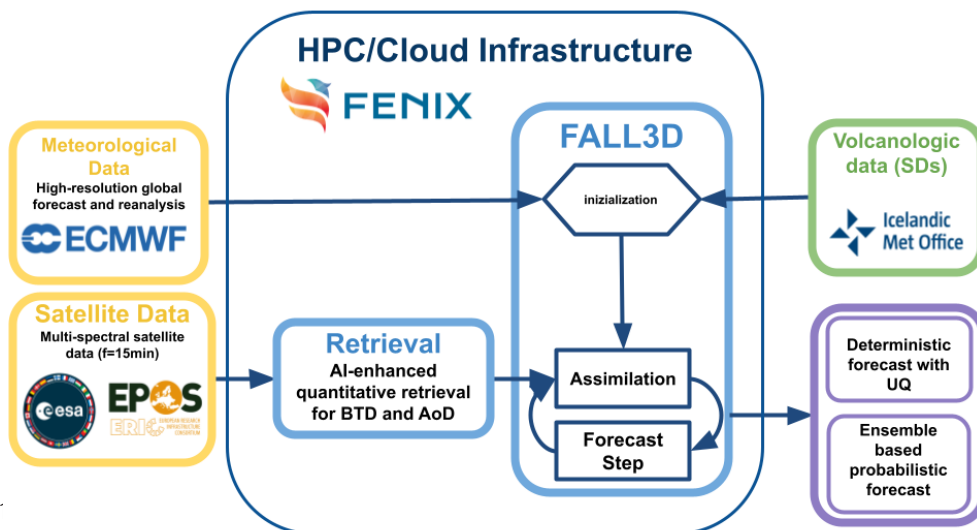


Fig. 4
 Sketch of the DTC for volcanic ash showing the data input sources and the assimilation and forecast loop process

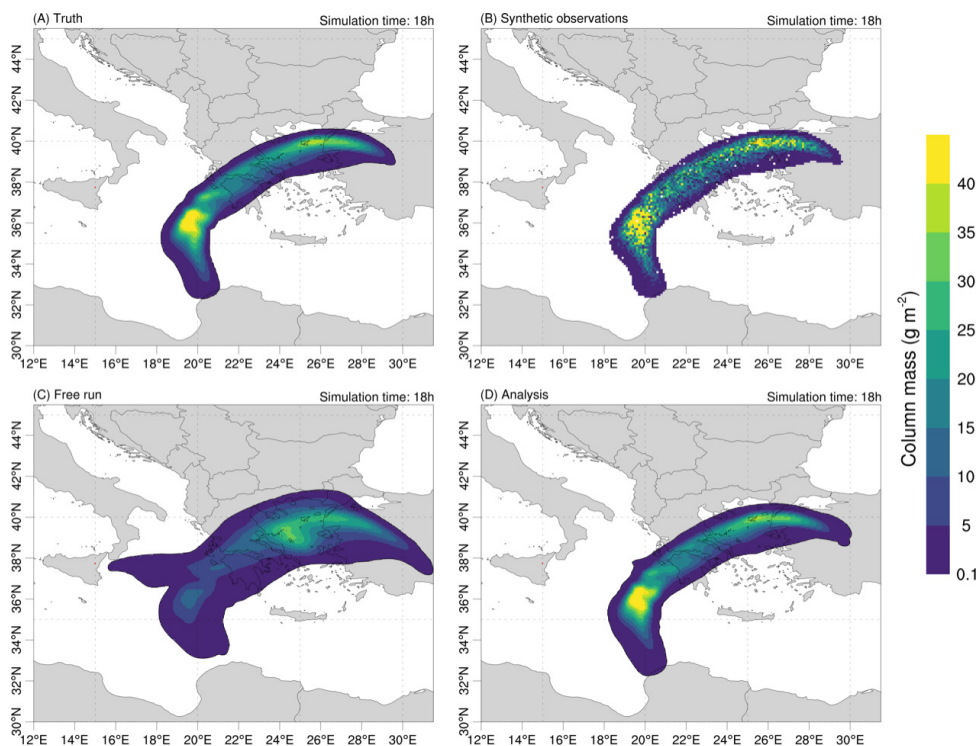


Fig. 5
Spatial distribution of ash mass loading for the DTC experiment at $t=18$ h after eruption start. From the true state (a) synthetic observations (b) are generated. Panel (c) shows the results from a run without assimilation, while in panel (d) the result of the DTC analysis with data assimilation are displayed (after Mingari et al. 2022)

4. Conclusions

This paper delves into the design guidelines for DT-GEO's data-driven workflows, with a spotlight on DTC implementation, and presents a case study on volcanic hazards. The case study provides a first example of the potential impact of DTCs/SDs in Geosciences. DT-GEO plans to test 3 additional SDs to consolidate these findings. However, the research team is prepared to switch to a different SD if a new eruption occurs, so as not to miss the chance to observe and monitor the evolution of the phenomenon in real time.

It's crucial to note that, while these events can have a significant impact on human activities, they are sporadic in nature. Capturing their behavior as it happens is crucial for knowledge advancement. If such observations are not possible, we can still rely on DTCs/SDs, which then become the only virtual laboratory available to us.

Finally, we plan to access the pre-exascale supercomputer Leonardo, hosted by the CINECA consortium at the Big Data Technopole in Bologna, Italy, and owned by EuroHPC JU. This will empower us to run interactive and massively scalable GPU-enabled workflows (Turisini et al. 2023).

Acknowledgement

We acknowledge the Project DT-GEO: Digital Twin for GEophysical extremes - funded by Horizon Europe under the grant agreement No 101058129, <https://cordis.europa.eu/project/id/101058129>

Bibliography

Alam, S. R., Bartolome, J., Carpene, M., Happonen, K., Lafoucriere, J. C., Pleiter, D. (2022), Fenix: a Pan-European federation of supercomputing and cloud e-infrastructure services. *Commun. ACM*, 65(4), pp 46-47.

Aubry, T. J., Engwell, S. L., Bonadonna, C., Carazzo, Scollo S., Van Eaton, A. R., Taylor I.A., Jessop D., Eychenne J., Gouhier M., Mastin L.G., Wallace K.L., Biass S., Bursik M., Grainger R.G., Jellinek M., Schmidt A. (2023), The Independent Volcanic Eruption Source Parameter Archive (IVESPA, version 1.0): A new observational database to support explosive eruptive column model validation and development, *J. Volcanol. Geoth. Res.*, 417, 107295

Carbonell R., Folch A., Costa A., Orlecka-Sikora B., Lanucara P., Løvholt F., Macias J., Brunne S., Gabriel A.-A. Barsotti, S., Behrens, J., Gomes J., Schmittbuhl J., Freda C., Kocot J., Giardini D., Afanasiev M., Galves H., Badia R. (2023), Digital Twinning of Geophysical Extreme Phenomena (DT-GEO), EGU General Assembly 2023, Vienna, Austria, 24–28 Apr 2023, EGU23-5674

Cocco, M., Freda, C., Atakan, K., Bailo, D., Contell, K. S., Lange, O., Michalek, J. (2022). The EPOS Research Infrastructure: a federated approach to integrate solid Earth science data and services. *Ann. Geoph.*, 65(2), DM208.

Costa, A., Macedonio, G., and Folch, A. (2006), A three-dimensional Eulerian model for transport and deposition of volcanic ashes, *Earth Planet. Sc. Lett.*, 241, pp. 634–647

Ejarque J. , Badia R. M. , Albertin L. , Aloisio G. , Baglione E. , Becerra Y. , Boschert S. , Berlin J. R. D’Anca A. , Elia D. , Exertier F. , Fiore S. , Flich, J, Folch A., ..., Volpe, M. (2022), Enabling dynamic and intelligent workflows for HPC, data analytics, and AI convergence, *Future Generation Computer Systems*, 134, pp. 414-429

Folch, A., Costa, A., Macedonio, G. (2009), FALL3D: A Computational Model for Transport and Deposition of Volcanic Ash, *Comput. Geosci.*, 35, pp. 1334–1342

Folch A., Abril C. , Afanasiev M., Amati G., Bader M., Badia R.M., Bayraktar H.B., Barsotti S., Basili R., Bernardi F, Boehm C., Brizuela B., Brogi F, Cabrera E., Casarotti E., Castro M.J., Cerminara M., Cirella A., Cheptsov A., Conejero J., ..., Wössner U., (2023), The EU Center of Excellence for Exascale in Solid Earth (ChEESE): Implementation, results, and roadmap for the second phase, *Future Generation Computer Systems*, 146, pp. 47-61,

Fuller A., Fan Z., Day C., Barlow C. (2020), Digital twin: Enabling technologies, challenges and open research. *IEEE access*, 8, pp. 108952-108971.

Gomes J., Bagnaschi E., Campos I., David M., Alves L., Martins J., ... Orviz P. (2018), Enabling rootless Linux Containers in multi-user environments: the udocker tool, *Comput Phys Commun.*, 232, pp. 84-97.

Global Volcanism Program (2019), Report on Raikoke (Russia), in Crafford, A.E., Venzke, E., (eds.), *Bulletin of the Global Volcanism Network*, 44 (8), Smithsonian Institution (USA)

Mingari L., Folch A., Prata A. T., Pardini F., Macedonio G., Costa A. (2022), Data assimilation of volcanic aerosol observations using FALL3D+PDAF. *Atmospheric Chem. Phys.*, 22(3), pp 1773-1792.

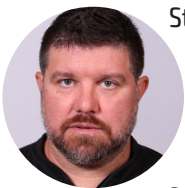
Talia D., Trunfio P. (2023), Urgent Computing for Protecting People From Natural Disasters, *Computer*, 56 (04), pp. 131-134.

Tejedor E., Becerra Y., Alomar G., Queralt A., Badia R. M., Torres J., ... Labarta, J. (2017), PyCOMPSs: Parallel computational workflows in Python. *Int J High Perform Comput Appl.*, 31(1), pp. 66-82.

Turisini, M., Amati, G., Cestari, M. (2023). LEONARDO: A Pan-European Pre-Exascale Supercomputer for HPC and AI Applications, preprint arXiv:2307.16885.

Wilkinson MD., Dumontier M., Aalbersberg IJ., 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., ..., Mons B. (2016), The FAIR Guiding Principles for scientific data management and stewardship, *Sci. Data*, 3: 160018.

Autori



Stefano Cacciaguerra stefano.cacciaguerra@ingv.it

PhD in Computer Science, he serves as Primo Tecnologo at the Istituto Nazionale di Geofisica e Vulcanologia. He designs and implements ICT infrastructure like HPC clusters, cloud architecture and data banks in order to support geophysical and volcanological applications. He is involved in projects DT-GEO and ChEESE-2P. He represents INGV in Italian Computing and Data Infrastructure. He develops ICT solutions for the EMSO Western Ionian Facility and manages EMSO's ICT initiatives in Bologna.

Antonio Costa antonio.costa@ingv.it

He is Senior Researcher at INGV Bologna, where he was also Director (2019-2022). Got the PhD at the University of Bologna (2004). He was Researcher at University of Bristol (2005-2006), where he was also honorary researcher (2007-2010). He worked at University of Reading (2011-2012). He visited the University of Tokyo (2015) and Munich University (2019). For his scientific achievements was awarded the IAVCEI Wager Medalist in 2013 and from 2021 is Member of Academia Europaea.





Francesca Quareni francesca.quareni@ingv.it

Degree and PhD in Physics, she is presently Senior Researcher at the Istituto Nazionale di Geofisica e Vulcanologia, after been employed as Post-Doc Researcher in Geophysics at Arizona State Univ. (USA), and as Associate Professor in Physics of Volcanism at Osservatorio Vesuviano in Naples. From 2013 to 2019 she served as Director of Bologna Branch of INGV. At present, she is member of the Board of Directors of the BigData Association, and represents INGV in the EOSC Association.

Flavio Cannavò flavio.cannavo@ingv.it

He is a Researcher at INGV Osservatorio Etneo in Catania, Italy. He earned both his Computer Engineering degree and Ph.D. in Complex Systems Engineering from the University of Catania. Flavio's expertise revolves around ground deformation modeling in volcanic and tectonic regions. He employs advanced statistical and machine learning techniques to enhance the understanding of Earth's dynamics, making his contributions to the field of Earth sciences and disaster management.



Arnau Folch afolch@geo3bcn.csic.es

He is a research Professor at the Geociencias Barcelona (GEO3BCN-CSIC). Degree in Physics and PhD in Applied Mathematics, he authors 120+ scientific peer-reviewed publications and has participated in 40+ Spanish and European competitive research projects. Coordinator of the EuroHPC Center of Excellence for Exascale in Solid Earth (ChEESE) and the DT-GEO project for digital twins in geophysical extremes. Member of the Strategic Advisory Board of the Destination Earth initiative.



Giovanni Macedonio giovanni.macedonio@ingv.it

He is Research Director at INGV Naples, where he was also Director (2001-2017). He graduated in physics in 1984. He was Researcher at CNR in Pisa (1988-1998). Since 2018, he has been a member of the Grandi Rischi Commission, volcanic risk sector: the Commission of the Civil Protection Department to support decisions in the event of a volcanic crisis. For his scientific achievements was awarded the IAVCEI Wager Medalist in 1998 and from 2019 is Member of Academia Europaea.



Sara Barsotti sara@vedur.is

She is Coordinator of Volcanic Hazards at the Icelandic Meteorological Office and the leader of the State Volcano Observatory team. She obtained her Ph.D at the University of Bologna (2006) after a Ms.D. in Physics at University of Pisa (2002). In the period 2006-2013 she worked as Researcher on contract at INGV, Sezione di Pisa (Italy). Dr Barsotti is and has been involved in several European and International collaborations and projects.



La piattaforma IEMAP (Italian Energy Materials Acceleration Platform): Il ruolo dei dati aperti e condivisi

Claudio Ronchetti, Sergio Ferlito, Marco Puccini, Simone Giusepponi, Francesco Buonocore, Massimo Celino, Sara Marchio, Giovanni Ponti

ENEA - Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile
TERIN - Dipartimento Tecnologie Energetiche e Fonti Rinnovabili ICT - Divisione per lo Sviluppo di Sistemi per l'Informatica e l'ICT

Abstract. La piattaforma IEMAP è costituita da una infrastruttura di supercalcolo e una rete di laboratori e ha lo scopo di condividere dati per accelerare la ricerca su nuovi materiali per l'energia. I dati sperimentali e computazionali prodotti vengono raccolti in un DB dedicato presso l'infrastruttura di supercalcolo. Mediante un approccio multidisciplinare che sfrutta le tecnologie Big Data e AI, la condivisione ed analisi dei dati potranno accelerare la progettazione dei materiali più adatti per una data tecnologia energetica. Materiali per batterie, elettrolizzatori e fotovoltaico (tre aree di ricerca fondamentali per la transizione energetica) sono i primi casi d'uso della piattaforma, il cui motore è il supercomputer CRESCO sul quale le tecnologie HPC sono implementate per la gestione dei dati e lo sviluppo/porting di software per il modelling molecolare. Alla piattaforma partecipano vari laboratori sperimentali e computazionali dei quattro enti di ricerca aderenti: ENEA, CNR, RSE e IIT.

Keywords. Interoperabilità e accesso a dati e servizi, Servizi ICT a supporto delle infrastrutture di ricerca e istruzione, Utilizzo e condivisione di dati in infrastrutture di rete e servizi.

Introduzione

Mission Innovation (MI) è un'iniziativa di cooperazione multilaterale con lo scopo di accelerare i processi di innovazione delle tecnologie pulite attraverso l'impegno dei paesi aderenti, tra i quali l'Italia, a raddoppiare la quota pubblica degli investimenti per le attività di ricerca, sviluppo e innovazione delle tecnologie per la decarbonizzazione al fine di rendere l'energia pulita accessibile ai consumatori e di creare posti di lavoro verdi e opportunità commerciali. In abito italiano sono coinvolti i quattro enti di ricerca: ENEA, CNR, RSE e IIT. L'iniziativa si articola in tre aree tematiche Smart Grid, Idrogeno e Materiali avanzati per l'energia attraverso i progetti: MISSION (Microreti e sistemi smart, multivettore ed integrati per accelerare la transizione energetica), HYDROGEN DEMO VALLEY (infrastrutture polifunzionali per la sperimentazione e dimostrazione delle tecnologie dell'idrogeno) e IEMAP (Piattaforma Italiana Accelerata per i Materiali per l'Energia). In quest'ultimo

progetto, il Dipartimento TERIN con la divisione ICT è direttamente coinvolto.

Il progetto ha come finalità la realizzazione di una piattaforma digitale che attraverso la condivisione ed elaborazione di grandi quantità di dati sperimentali e computazionali favorisca la progettazione accelerata e la selezione dei materiali avanzati per l'energia. Piattaforma multidisciplinare che in modo automatico, combinando tecnologie BigData e di AI, possa accelerare il processo di analisi dei dati computazionali e sperimentali al fine di identificare i materiali più adatti per una data tecnologia energetica. Materiali per l'accumulo elettrochimico-batterie, per gli elettrolizzatori e per il fotovoltaico, rappresentano aree di ricerca centrali per favorire il processo di transizione energetica. Verranno perciò semplificati quei processi, complessi e costosi, necessari per il design di nuovi materiali che comprendono la simulazione, la sintesi e la caratterizzazione con l'esecuzione di numerosi test.

Il motore di questa piattaforma è l'infrastruttura ENEAGRID e in particolare il supercomputer CRESCO, installato presso il C. R. ENEA di Portici, su cui si implementeranno tecnologie HPC sia per la gestione dei dati sia per lo sviluppo ed implementazione di codici numerici per il modeling molecolare. Alla piattaforma partecipano vari laboratori sperimentali e computazionali dei quattro enti di ricerca aderenti al progetto.

1. Gestione dei dati

Nelle fasi iniziali del progetto, una sfida importante è stata quella di definire un modello di metadato comune per i domini sperimentale e computazionale provenienti dai differenti laboratori e simulazioni. A questo scopo si è scelto di basare la definizione del modello dati sull'ontologia riportata in (Andersen 2021). Il modello logico è rappresentato da quattro entità (Fig. 1) che costituiscono le informazioni del materiale preso in analisi; a) il processo utilizzato per analizzare il materiale, b) i parametri che incidono sul processo, c) le proprietà chimico-fisiche ricavate dal processo e d) la provenienza del dato generato. A questo possono essere legati file di dati grezzi generati dal processo.



Fig. 1
Rappresentazione schematica del metadato con le informazioni del materiale considerato: processo, parametri, proprietà e provenienza. Enti di ricerca coinvolti nel progetto

2. Piattaforma digitale

Al fine di raccogliere, gestire e immagazzinare dati eterogenei provenienti da contesti sperimentali e numerici, avendo come riferimento i principi FAIR (Findable, Accessible, Interoperable, Reusable) (Wilkinson 2016) è stata realizzata un'infrastruttura digitale (Fig. 2) composta da quattro elementi principali, ovvero:

1. Interfaccia Web (front-end);
2. REST API (Application programming interface);
3. NoSQL DataBase;
4. File Storage S3.

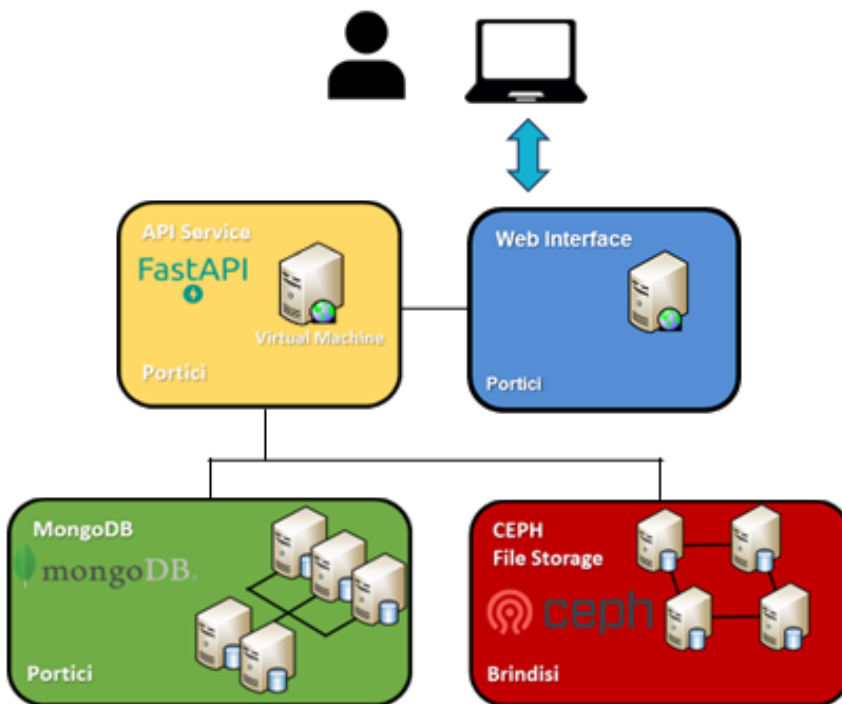


Fig. 2
Rappresentazione schematica della piattaforma digitale: Web Interface, API service, MongoDB e CEPH File Storage

2.1 Interfaccia Web

L'interfaccia Web [iemap.enea.it] permette di rendere fruibili le funzionalità implementate nelle API agli utenti persone tenendo conto degli aspetti di usabilità e accessibilità. VUE 3 è la tecnologia utilizzata per l'interfaccia Web.

2.2 REST API

Le REST API sono indispensabili per la creazione e l'integrazione di software applicativi.

In questo caso forniscono diverse funzionalità legate agli accessi, al caricamento dei dati e alle interrogazioni sul DB. Le API rappresentano uno strato di sicurezza superiore per l'accesso ai dati e, inoltre, permettono interazioni sia con utenti macchina (direttamente) sia con utenti persone (indirettamente via interfaccia Web). Le REST API si basano su tecnologia FastAPI.

2.3 DataBase

Il Database NoSQL permette di immagazzinare i metadati (formato JSON) caricati dagli utenti partner della piattaforma e di effettuare query elaborate in modo efficiente. MongoDB è la soluzione adottata come database.

2.4 File Storage

Il File Storage S3 permette di conservare i file grezzi ottenuti dai processi che sono strettamente legati al metadato. La tecnologia per l'immagazzinamento dei dati grezzi è quella fornita da CEPH.

3. Caso d'uso

Il nostro gruppo sta svolgendo ricerca sui materiali per l'accumulo elettrochimico. Lo scopo è quello di trovare alternative alle batterie a base di ioni di Litio (LIB) con altre a base di ioni di Sodio (NIB) vista la sempre più limitata disponibilità di Litio (Li) a fronte della crescente richiesta di mercato. Invece, il Sodio (Na), essendo il 6° elemento chimico per disponibilità, è poco costoso e ampiamente distribuito rendendo le NIB una valida alternativa alle LIB.

Le NIB sono batterie ricaricabili analoghe alle LIB. Il catodo è un materiale allo stato solido contenente gli ioni alcalini, l'anodo è un materiale allo stato solido non contenente necessariamente gli ioni alcalini e l'elettrolita è una soluzione contenente gli ioni alcalini. In entrambi i catodi viene utilizzato il meccanismo di reazione di intercalazione. A causa dei loro elevati potenziali operativi e delle elevate capacità, i catodi a base di sodio hanno ricevuto crescente attenzione. Perciò, buone prestazioni e costi inferiori, insieme al fatto di poter utilizzare la stessa tecnologia e produzione già sviluppata per le LIB senza costi aggiuntivi, hanno suscitato un rinnovato interesse per questi sistemi.

Gli svantaggi al momento riguardano la densità di energia e la ciclabilità: nelle NIB la densità di energia è ancora inferiore rispetto alle LIB e la struttura geometrica può subire una distorsione Jahn Teller durante il ciclo di scarica/ricarica. Molti studi sono dedicati al miglioramento delle prestazioni delle NIB.

In particolare, si è visto che l'ossido metallico lamellare NaMnO_2 in geometria P2 opportunamente drogato, mostra buone proprietà di ciclicità e capacità (Lee 2013, Wang 2013, Prosini 2019). In questo ambito si è definito un protocollo di calcolo ab-initio delle energie di formazione e dei potenziali redox (la prima è legata alla stabilità del composto e quindi alla ciclabilità, il secondo è legato alla densità di energia) e, partendo dal cristallo base, è stato predisposto un workflow per la generazione automatica di modelli atomistici di NaMnO_2 drogato cui si sono variati l'elemento (Ni e Ti) e il livello di drogaggio. Viste

le numerose combinazioni possibili e nonostante le sempre maggiori capacità di calcolo, un approccio basato sui soli codici quantistici risulta proibitivo. Si è fatto perciò ricorso a tecniche di IA per fare delle previsioni sui tipi di strutture più promettenti andandone a prevedere l'energia di formazione e il potenziale redox. Dall'analisi si è visto che queste sono quelle che hanno concentrazioni simili dei tre elementi (Mn, Ti, e Ni) e con distribuzioni spaziali il più possibile uniformi (Ronchetti 2022). Indicazioni confermate da calcoli con codici quantistici a livello atomistico.

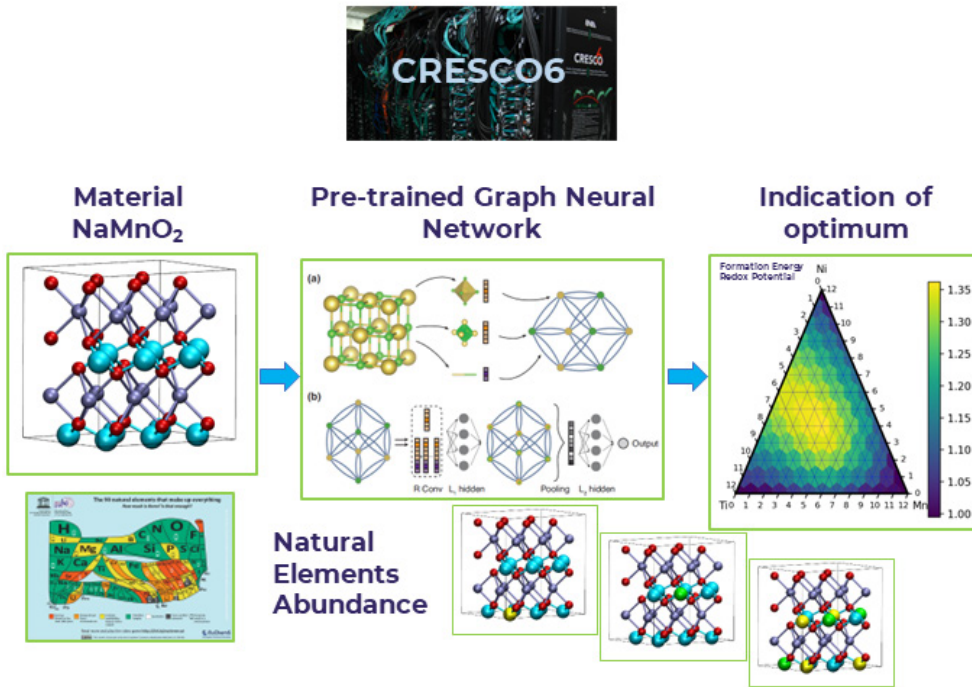


Fig. 3
Rappresentazione del workflow che dal materiale di partenza NaMnO_2 , attraverso tecniche di AI predice le concentrazioni droganti ottimali di Ni e Ti.

4. Conclusioni

La piattaforma IEMAP è costituita da una infrastruttura di supercalcolo e una rete di laboratori afferenti ai quattro enti partecipanti all'iniziativa MI con lo scopo di condividere dati per accelerare la ricerca su nuovi materiali per l'energia. La piattaforma digitale, costituita da quattro componenti principali che sono l'interfaccia Web, le API, il DB e il File Storage, permette la condivisione dei dati sperimentali e computazionali e, mediante un approccio multidisciplinare potrà accelerare la progettazione di materiali avanzati per le batterie, gli elettrolizzatori e il fotovoltaico.

Ringraziamenti

Questo lavoro fa parte del progetto IEMAP finanziato dal MASE (Ministero dell'Ambiente e della Sicurezza Energetica) nell'ambito dell'iniziativa internazionale MI (Aspuru-Guzik 2018).

Riferimenti bibliografici

Andersen C. W. et al. (2021), OPTIMADE, an API for exchanging materials data, *Sci. Data*, (8), 217.

Aspuru-Guzik A. et al. (2018), Materials Acceleration Platform: Accelerating Advanced Energy Materials Discovery by Integrating High-Throughput Methods and Artificial Intelligence, *Mission Innovation, Innovation Challenge 6*.

Lee D. H. et al. (2013), An advanced cathode for Na-ion batteries with high rate and excellent structural stability, *Phys. Chem. Chem. Phys.*, (15), pp. 3304-3312.

Prosini P. P. et al. (2019), Tin-Decorated Reduced Graphene Oxide and NaLi_{0.2}Ni_{0.25}Mn_{0.75}O as Electrode Materials for Sodium-Ion Batteries, *Materials*, (12), 1074.

Ronchetti C. et al. (2022), Machine Learning Techniques for Data Analysis in Materials Science. 2022 AEIT International Annual Conference (AEIT), Rome, Italy, 2022, pp. 1-6.

Wang, H. et al. (2013), Electrochemical properties of P2-Na_{2/3}[Ni_{1/3}Mn_{2/3}]O₂ cathode material for sodium ion batteries when cycled in different voltage ranges, *Electrochim. Acta*, (113), pp. 200-204.

Wilkinson M. D. et al. (2016), The FAIR Guiding Principles for scientific data management and stewardship, *Sci. Data*, (3) 160018.

Autori



Claudio Ronchetti claudio.ronchetti3@gmail.com

Claudio Ronchetti ha conseguito la laurea in Ingegneria Informatica presso l'Università di Roma Tre. Nel corso del suo percorso professionale ha applicato soluzioni di AI all'avanguardia in ambito pubblico e privato, spaziando in vari domini come, ad esempio, il rilevamento delle frodi, la scienza dei materiali, la trascrizione di documenti storici e digitali, la computer vision e l'elaborazione del linguaggio naturale.

Sergio Ferlito sergio.ferlito@enea.it

Sergio Ferlito è nato a Napoli, Italia nel 1969. Laureato in Ingegneria Elettronica presso l'Università degli Studi di Napoli "Federico II" nel 1998.

Dal 1999 lavora presso l'ENEA, Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile. I suoi interessi di ricerca includono modelli statistici e di machine learning applicati ai settori delle energie rinnovabili e dell'IoT, sviluppo software web (front/back-end), framework di machine learning online e DevOps.



Marco Puccini marcopuccini@pm.me

Laureato in Fisica all'Università di Roma "La Sapienza", ha lavorato con l'INM del CNR, il

DIMA ed il Dipartimento di Medicina Molecolare de "La Sapienza" come data scientist. Come Assegnista di Ricerca presso la Divisione ICT del Dipartimento TERIN di ENEA si è occupato principalmente di data management, sviluppo di soluzioni software cloud in diversi progetti come Mission Innovation IEMAP, DTECH, ECODIGIT, RAFAEL.



Simone Giusepponi simone.giusepponi@enea.it

Ricercatore presso la Divisione ICT del Dipartimento TERIN dell'ENEA. Dottorato in Fisica con una tesi nel campo delle simulazioni numeriche applicate ai sistemi non lineari. I suoi interessi di ricerca riguardano le applicazioni del calcolo parallelo e ad alte prestazioni con esperienze di ricerca nell'area della scienza dei materiali. Coinvolto in progetti di ricerca italiani ed europei come Mission Innovation IEMAP, VIPERLAB, StoRIES, ICSC, EoCoE-II, I-NEST, ROMA-TECHNOPOLE.

Francesco Buonocore francesco.buonocore@enea.it

Ricercatore ENEA dal 2012, lavora nella Divisione ICT del Dipartimento TERIN. Fisico specializzato in teoria della materia condensata, ha conseguito il dottorato presso l'Università di Napoli Federico II nel 2001. Le sue ricerche si concentrano su calcoli ab initio per nanomateriali e interfacce. È impegnato in progetti di ricerca italiani ed europei quali Mission Innovation IEMAP, StoRIES, ICSC, EoCoE e I-NEST ed è co-inventore di diversi brevetti e autore di numerose pubblicazioni.



Massimo Celino massimo.celino@enea.it

biografia: Massimo Celino, PhD in Fisica della Materia (Università di Strasburgo, Francia), laureato in Fisica (Università "La Sapienza" di Roma). Interessi: modellistica molecolare, calcolo ad alte prestazioni, gestione dei dati. Coinvolto in progetti europei (EoCoE, EERA-data), vicecoordinatore JP EERA "Digitalization for Energy", coordinatore progetto EuroHPC "Textarossa" per tecnologie exascale. 90+ pubblicazioni scientifiche.

Sara Marchio sara.marchio@enea.it

biografia: Ricercatrice ENEA presso la Divisione ICT del Dipartimento TERIN. Laurea in Fisica nel 2015 e PhD in Meccanica Teorica e Applicata nel 2019 (Università di Roma "La Sapienza") con una tesi nel campo delle simulazioni atomistiche combinate a tecniche per eventi rari con applicazioni nell'ambito della fluidodinamica. I suoi interessi attuali sono relativi alla scienza dei materiali ed al calcolo ad alte prestazioni. È coinvolta nei progetti ICSC e ROME-TECHNOPOLE.



Giovanni Ponti giovanni.ponti@enea.it

biografia: Direttore della Divisione ICT dell'ENEA e membro del CTS del GARR. PhD in Ingegneria Informatica nel 2010 (Università della Calabria) concentrandosi sulla modellazione dei dati e gli algoritmi di data mining. Ricercatore ENEA per 12 anni con attività di ricerca nell'HPC, Data Science, Cloud Computing, Big Data e Data Analytics. Campi nei quali è stato autore di prestigiose pubblicazioni scientifiche, membro del comitato di programma di conferenze e revisore per importanti riviste e convegni.

The RESTART project: RESearch and innovation on future Telecommunications systems and networks, to make Italy more “smART”

Cristina E. Costa¹, Antonella Bogon¹, Ilenia Tinnirello², Guido Maier³, Giacomo Verticale³, Raffaele Bolla⁴

¹CNIT - Consorzio Nazionale Interuniversitario per le Telecomunicazioni,

²Università degli Studi di Palermo, ³Politecnico di Milano, ⁴ Università di Genova

Abstract. RESTART is an extended partnership program kicked-off on January 2023, funded by the European Union – NextGenerationEU. This comprehensive program targets at aggregating, and integrating the efforts of the main stakeholders of the Italian Telecommunications sector, in various specific research projects positioned in an overall framework providing a long-term vision, transversal and supporting actions, and adequate enablers. The cooperation between universities, research centres, companies, and public administrations, shall overcome the current fragmentation of research and development through actions of coordination and synergy between projects. To achieve concrete and measurable results, all phases, from ideas and research (with low TRL) to technological innovation and services (with attention to possible opening with high TRL) are addressed, as well as the birth and growth of companies, transforming research results into practical innovation and value creation for the Italian telecommunications ecosystem. We expect RESTART to produce long-lasting and sustainable results beyond the duration of the program

Keywords. Telecommunications, NextGenerationEU, Technological enablers, Innovation, Sustainability

Introduction

We are entering in the Post-Information Society, fueled by a plethora of various emerging smart technologies enormously enhanced by a never seen before communication capability. Innovative applications and services are flourishing in the most diverse sectors (agriculture, commerce, energy, finance, manufacturing/industry, media, health, safety/security, transportation), changing relationships and distributing intelligence. In this scenario, the telecommunications infrastructure is already playing a fundamental role in interconnecting the human, the digital, and the physical worlds, and setting the premises for becoming in the future, even more than today, a fundamental Critical Infrastructure (CI) of our modern societies, since communication is now the fabric of our societies.

To cope with the increasing demands of the Post-Information Society, future infrastructure shall target not only performance improvement of key indicators, such as service availability, energy efficiency, data rates, etc., but also new societal key values indicators (KVI) such as trustworthiness, inclusiveness, and sustainability addressing human and

societal needs (6G IA). To effectively design this future critical infrastructure, new approaches and paradigms are needed from academia and industry, leveraging on synergy and multidisciplinary.

In this context, RESTART is conceived as a comprehensive program, targeting at aggregating, and integrating the efforts of the main stakeholders of the Italian Telecommunications sector (Fig. 1), in various specific research projects positioned in an overall framework that provides a long-term vision, transversal and supporting actions and adequate tools and enablers (Restart).

Università di Roma, Tor Vergata
Politecnico di Bari
Politecnico di Milano
Politecnico di Torino
Università degli Studi di Catania
Università degli Studi di Firenze
Alma Mater Studiorum – Università di Bologna
Università degli Studi di Napoli "Federico II"
Università degli Studi di Padova
Scuola Superiore Sant'Anna
Università degli Studi Mediterranea di Reggio Calabria
Università degli Studi di Roma "La Sapienza"
Consorzio Nazionale Interuniver. per le TLC
Consiglio Nazionale delle Ricerche
Fondazione Ugo Bordoni
Open Fiber S.p.A.
Prysmian S.p.A.
Athonet S.R.L.
Tiesse S.p.A.
Wind Tre S.p.A.
Vodafone Italia S.p.A.
Tim Italia S.p.A.
Italtel S.p.A.
Ericsson Telecomunicazioni S.p.A.
Leonardo S.p.A.

Fig. 1
List of current RESTART partners. The color identifies the type, i.e. universities, national research organizations and private organizations. Partners which are also Spoke coordinator are in bold

The strength of this program is the cooperation between universities, research centres, companies and public administrations, that will allow overcoming the current fragmentation of research and development through actions of coordination and synergy between project initiatives. To achieve concrete and measurable results, a specific effort has been put in identifying actions that cover all phases from ideas and research (with low TRL) to technological innovation and services (with attention to possible opening with high TRL), and then to the birth and growth of companies, transforming research results into practical innovation with higher TRL and value creation for the Italian telecommunications ecosystem. The final goal will be being able to produce long-lasting and sustainable results

beyond the duration of the program.

Besides its research and innovation goals, RESTART has as third pillar also social responsibility and education goals, for contributing to society grand challenges related to climate change, circular economy and smart infrastructures, as well as training a new generation of researchers and innovators able to bring the future of telecommunications on their shoulders.

Research projects in RESTART are of two types: (1) structural projects which define the program, they are the backbone of the partnership and give a structure to the overall research work, focusing on the cooperation of a significant number of partners, and integrating the results of a rather large research area. And (2) focused projects which are smaller endeavors, with fewer researchers and partners, aimed at completing the structural projects in specific directions. Focused projects can be of various nature: industrial, targeting specific requirements and needs coming from industrial partners or targeting a definite application scenario; theoretical focusing on long-term research frontiers of special interest; or blue-sky research, not necessarily referred to one or more structural projects.

1. Research Infrastructures role in RESTART

The RESTART program, is designed around 7 missions, that complement the Research Mission (Fig. 2).

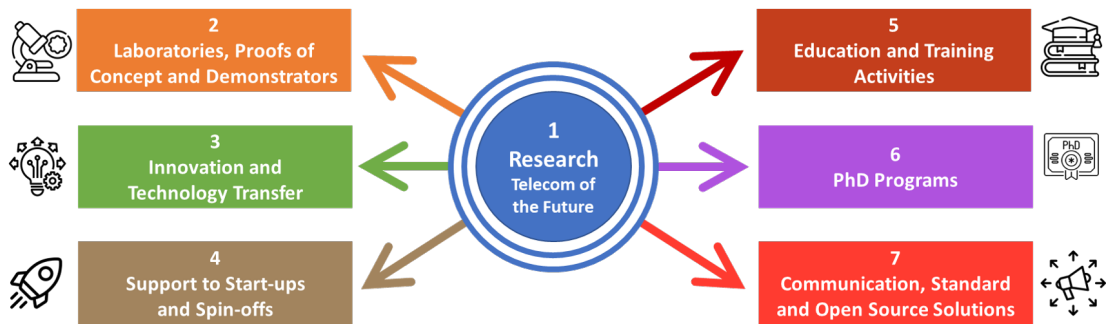


Fig. 2

General scheme of the RESTART program, displaying all the project missions

One of the objectives of RESTART is to boost up the current Italian Research Infrastructure for the Telecommunications by enlarging the existing laboratories and creating new, shared, ones: the whole Mission 2 is dedicated to “Laboratories, Proofs of Concept and Demonstrators”.

Creating a solid Telecommunications Research Infrastructure (RI) is a key action for long-term impact of the program beyond the projects’ lifespan. The RI in RESTART supports the effectiveness of the research and training missions, and also works towards the consolidation of collaborations. The final ambition is the creation of an Italian ecosystem of telecommunications laboratories and testing facilities where industrial and academic

research can have a shared playground able to stimulate fruitful collaboration and cross-fertilization at different TRLs.

The strategy for achieving this goal includes: i) setting up shared experimental labs and testing platforms tailored to the specific domains; ii) expanding and extending existing laboratories and developing open testbeds at the partners' premises; iii) establishing access rules of both academic and industrial labs that ensure collaboration in the shared facilities; iv) planning for medium-and long-term consolidation and extension of the experimental instruments with other funding sources; v) enhancing in particular the experimental facilities located in the South of Italy.

All resources will be consolidated according to the needs of research institutions and companies to cope with the new technologies, including new higher frequency bands for future wireless network deployments. In addition, new platforms oriented to application scenarios and domain specific networks will be created for research and innovation activities in new emerging research fields such as fully immersive communications, secure communications, convergence of sensors and communications, connectivity in challenging environments, artificial intelligence applications, and others.

All the mentioned infrastructures and laboratories will be exploited for carrying out the proofs of the concept proposed in RESTART.

2. The role of RIs in the evolution of Networking technologies for a Future sustainable and inclusive society

Net4Future is one of the RESTART's Structural projects, and it has the specific ambition of putting in first place society-pull requirements for future digital services (Hexa) by identifying technological enablers and innovation potential. Inclusiveness (to connect the unconnected) and sustainability (to contribute to sustainable utilization of natural resources in different sectors), trustworthiness (to control technology at different user/operator/provider levels), innovative regulation policies (including pricing regulation, competition regulation, privacy and data protection, online dispute resolutions) are the four essential directions to be considered while designing future telecommunications critical infrastructure. Net4Future makes a synthesis of all the relevant outputs of the 18 projects of the RESTART program and offers an experimental toolbox for validating innovative services.

In this context Research infrastructure plays the role of providing tools for validating solutions for future networks, building an end-to-end experimental testbed of future digital services, integrating heterogeneous technology enablers and components, and monitoring performance indicators. Net4Future supports the research in the domain of future Internet networks and beyond-5G (B5G) systems to demonstrate trade-offs across topics investigated in other projects (energy vs AI/ML vs efficiency vs usability). The objective also deals with monitoring solutions and the creation of data repositories.

3. From national to international

In an effort to expand the approach to an International EU level, RESTART RIs closely col-

laborate with the ESFRI Scientific Large-scale Infrastructure for Computing/Communication Experimental Studies (SLICES), a distributed flexible Digital Infrastructure designed to support large-scale, experimental research focused on networking protocols, radio technologies, services, data collection, parallel and distributed computing and in particular cloud and edge-based computing architectures and services (ESFRI). The participation of two RESTART partners, CNIT and CNR, with their distributed facilities to both the projects, assures the capability and potential to have an impact beyond the national stage.

Acknowledgement

This work was partially supported by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, partnership on “Telecommunications of the Future” (PE00000001 - program “RESTART”).

References

6G IA, (2022) Vision and Societal Challenges WG, Societal Needs and Value Creation Sub-Group. “What societal values will 6G address? Societal Key Values and Key Value Indicators analysed through 6G use cases,” White Paper

Hexa-X D1.2 “Expanded 6G vision, use cases and societal values”, <https://hexa-x.eu/deliverables/> (accessed 12/04/2022)

ESFRI SLICES Web Sites <https://slices-sc.eu/>

Restart Web Site <https://www.fondazione-restart.it/>

Authors



Cristina E. Costa cristina.costa@ieee.org

Cristina E. Costa is a Senior Researcher at the National Laboratory on Smart and Secure Networks at CNIT, the National, Inter-University Consortium for Telecommunications. She gained experience in the fields of wireless and mobile networks, multimedia communications, interfaces and interaction, 5G, and edge computing working in various research centers (CSELT, CREATE-NET, FBK). She was involved in several research projects both at the national and international level and served in the organizing committees of various conferences, as European Wireless, Intetain, UCMedia, and SecureComm. She is a founding member and past secretary of the IEEE Women In Engineering AG Italy Section.

Antonella Bogoni antonella.bogoni@cnit.it

Antonella Bogoni, full Professor at Sant'Anna School, is one of the pioneers of the Integrated Research Center for Photonic Networks and Technologies created in Pisa in 2001 by Sant'Anna School and CNIT where currently she is director of the National Laboratory on Photonic Networks and Technologies-PNTLab and leader of the “digital & microwave photonics” area. Antonella Bogoni dedicated her research activity to photonics technologies for optical communication and sensing.





Ilenia Tinnirello ilenia.tinnirello@unipa.it

Ilenia Tinnirello is a Full Professor of Telecommunications at the Department of Engineering of the University of Palermo. Her main research interests are focused on emerging technologies for mobile radio networks, and in particular on low-power long-range technologies, programmable cellular networks and wireless testbeds. She is currently a member of the Board of Directors of the National Interuniversity Telecommunications Consortium (CNIT) and the scientific coordinator of the European Digital Innovation Hub called i-NEST. Within the RESTART program, she is the PI of the Net4Future structural project.

Guido Maier guido.maier@polimi.it

Guido Maier received his Laurea (1995) and his Ph.D. degree (2000) at Politecnico di Milano (Italy). Until 2006 he has been with CoreCom (research consortium supported by Pirelli in Milan, Italy). On 2006 he joined the Politecnico di Milano as Assistant Professor. In 2015 he became Associate Professor. He is author of more than 200 papers (h-index 28) and 6 patents. He is PI of the RESTART Focused Project WatchEDGE. He is a Senior Member of the IEEE Communications Society.



Giacomo Verticale giacomo.verticale@polimi.it

Giacomo Verticale received the Ph.D. degree in telecommunications engineering from the Politecnico di Milano, Italy, in 2003. He is currently an Associate Professor at the Politecnico di Milano. He was involved in several research projects on fixed and wireless broadband access technologies and promoting the smart grid. His current interests focus on the security issues of the smart grid, on network function virtualization, and on edge computing in 5G. He is PI of the RESTART Focused Project LEGGERO.

Raffaele Bolla raffaele.bolla@unige.it

Raffaele Bolla is a Full Professor of Telecommunications Networks at the Department of Naval, Electrical, Electronic, and Telecommunications Engineering (DITEN) of the University of Genoa. He is currently Vice-President and a member of the Board of Directors of the National Interuniversity Telecommunications Consortium (CNIT), and he is the Director of the Smart and Secure Networks (S2N) CNIT National Laboratory. He has been and is responsible-Principal Investigator for many research projects, primarily in the EU contest. He is also active in standardization, mainly in ETSI and ITU-T. Currently, he is responsible for CNIT's participation in the national project RESTART



The IT@CHA Virtual Lab: a set of enabling technologies for Digital Twin implementation and enrichment in Cultural Heritage

Marialuisa Mongelli, Beatrice Calosso, Samuele Pierattini, Marco Puccini, Giovanni Ponti

ENEA

Abstract. The ongoing digital revolution is outlining efficient ways of integrating Digital Twin, 3D computing graphics, and structural behavior simulations in all the governance phasis of cultural assets with the final goal of developing new models for conservation and end-users' engagement. Thus, the platform IT@CHA (Italian Technologies for advanced applications on CH), has been updated since develops the virtual Lab: a multimedia laboratory embedded in the Computational RESsearch Centre on COMplex systems, CRESCO. The paper shows the capabilities of the IT@CHA Virtual Lab about the entire numerical modelling chain, useful on cultural assets, according to FAIR principles focused at highlighting its interoperability, scalability, and versatility. Users of the Virtual Lab can develop numerical models and manage 3D data to support the preservation phase of the asset.

Keywords. Digital Twin, data interoperability, Open Science & FAIR Data, CH enhancement

Introduzione

CRESCO (<https://www.eneagrid.enea.it/CRESCOportal/>) is the High-Performance Computing (HPC) facility developed by ENEA, with remote web access, able to share HPC resources, structural computation, and graphics resources. It is a complex infrastructure based on a cluster of 434 nodes (servers) each with top level hardware resources that brings to 20832 cores. All those resources are available through a remote system based on the SSH protocol that allows users to work with. That infrastructure has a unique authentication system and with both the General Parallel File System (GPFS, brand name IBM Spectrum Scale) and Andrew File System (AFS) makes it possible for users to have a personal and a shared environment to work with. In particular, the GPFS is a high-performance clustered file system, optimized for huge workload such as the required for HPC that can be used like a shared area; whereas AFS is a distributed file system that allows to have the same own folders on all the nodes, like what a common user has in his own laptop or desktop device.

On top of that hardware resources some Virtual Labs (VL) are being developed during past project work packages: they are a set of software tools that use CRESCO resources, served remotely via web for researchers' activities. There are different VL for different research areas of interest. Here we present two of them: DySCo (Structural Dynamic, numerical Simulation qualification tests and vibration Control) and IT@CHA (Italian Technologies for advanced applications on Cultural Heritage): <http://www.laboratorivirtuali.enea.it/it/>

prime-pagine/it-cha-virtual-lab.

1. IT@CHA Virtual Lab

The IT@CHA Virtual Lab (Figure1) is a digital platform developed for the IT@CHA project started in 2013 by ENEA (<http://www.progettoitacha.it/>), and still available today, with continuous updates and new features implementation. It is fully integrated in the ENE-AGRID/CRESCO, where the main High-Performance Computing cluster is hosted in the ENEA Portici Research Centre. Such a platform is reachable via web thanks to dedicated servers that share graphics resources on a high-performance network, allowing the usage of professional software both proprietary and open source, for computer vision workflows. Thanks to a simplified user interface it is possible to work remotely with tools like Agisoft Photoscan, Meshlab, Nastran/Patran or Blender all inside the same environment, enforcing interoperability and integration. Storage tools like the distributed file systems AFS and IBM GPFS allow data sharing among users and software, while a novel system like E3S (Enea Sharing Staging Storage) allow to load experimental and/or computational outputs into a unique area to make it available for researchers. All these software take advantage of CRESCO computing resources that can accelerate performance.

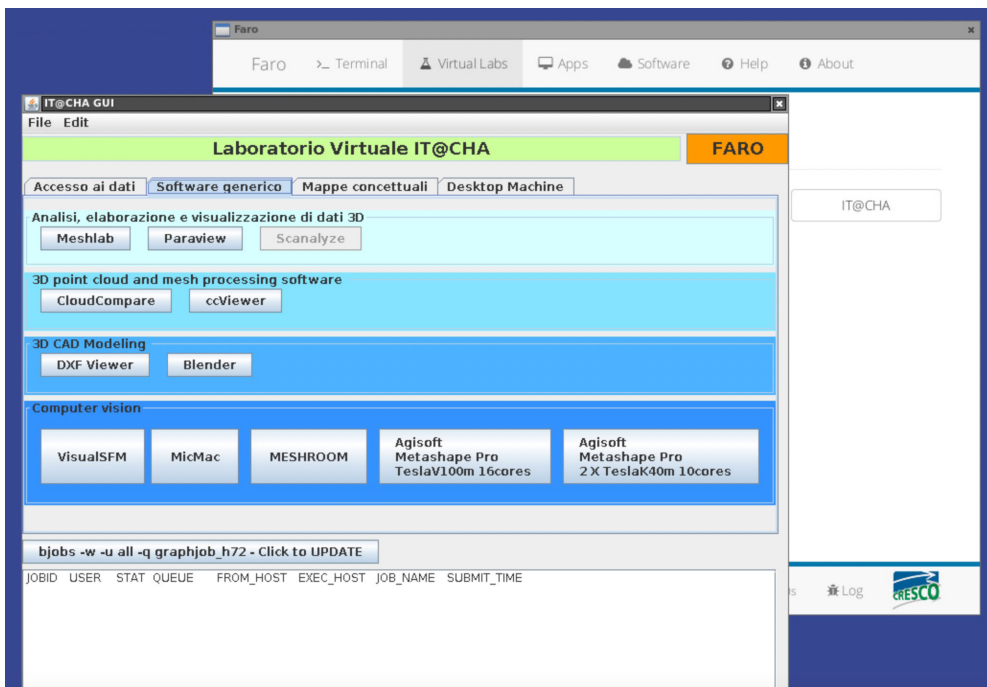


Fig. 1
The interface of
the IT@CHA Virtual
Lab

New implementations and features have been developed in the last few years, such as a dedicated modern NoSQL database to archive metadata of digital objects, making it available for further works and research and to easily integrate with advanced viewer services like 3D HOP for 3D objects visualization or IIPImage server for Gigapixel High resolution

images. Such new features are hosted on a dedicated virtual machine (heritagescience.portici.enea.it) managed into a VMWare cluster to ensure his redundancy and backup. In the last five years, starting from the experience of IT@CHA and the collaborations within the Lazio Region Culture Technology District Center of Excellence (DTC Lazio - <https://dtclazio.it/>), the ENEA Digital Twin Environment for Cultural Heritage was implemented also thanks to D-TECH (Digital Twins for Cultural Heritage) project that is going on, with the aim of building a distributed infrastructure for Digital Twin management for cultural assets, adopting the Data Mesh paradigm to leave control over their data as much as possible in the hands of heritage managers.

2. Use case

The use case of the “Bridge of the Towers” in Spoleto shows features and potentiality of IT@CHA platform performed within the Resilience Enhancement of a Metropolitan Area Project (RoMA). The survey aimed at defining the state of conservation of the bridge and at ensuring its long-term preservation using numerical models, starting from a drone stereo-photogrammetric scanning to Finite Element (FE) modelling and many other monitoring techniques, to assess the bridge cracks pattern - without building scaffoldings - and its structural health by a multidisciplinary approach that allows their mutual validation. The scanning produced 818 HD images (38 Mpix), respectively 411 for the north side and 407 for the south one with a GSD (Ground Sample Distance) equal to 1.75 mm/pixel. Thanks to a GPS module embedded in the drone, the photos were also completely georeferenced. The images acquired during the drone flights were post-processed through the ENEA HPC resources provided by CRESCO by the Structure from Motion (SfM) technique, which allows to estimate and reconstruct 3D structures from two-dimensional images. The technique performs photogrammetric processing of digital images and generates 3D spatial data and a dense point cloud useful for visual effects production and indirect measurements of the objects (Figure 2-3-4).



Fig. 2
Dense cloud of
the North side of
the Bridge of the
Towers in Spoleto



Fig. 3
Dense cloud of
the South side of
the Bridge of the
Towers in Spoleto.

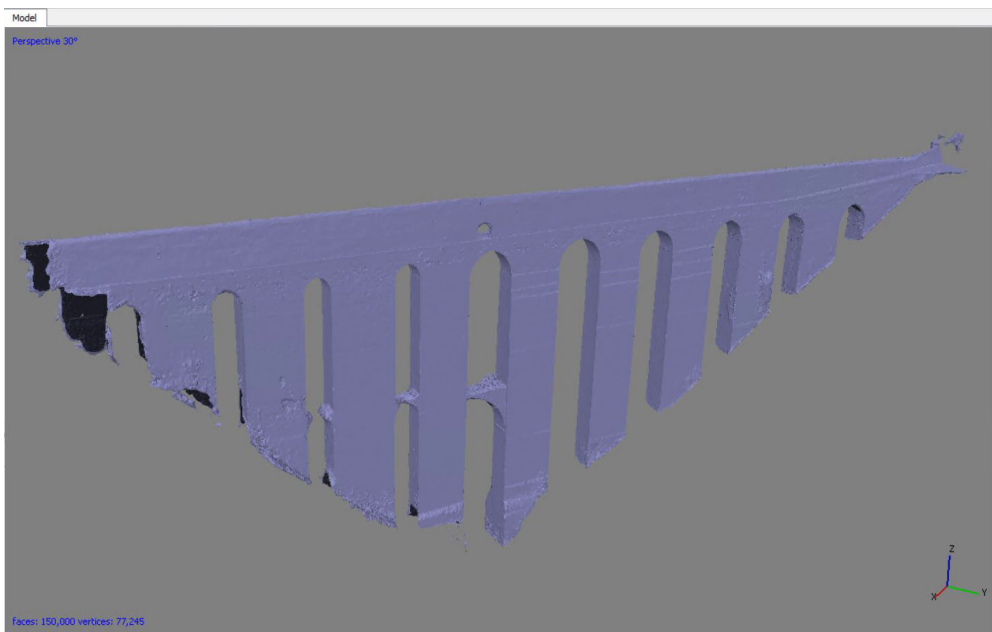


Fig. 4
Mesh of the Brid-
ge of the Towers
in Spoleto

The preliminary 3D FE model (Figure 5), performed based on geometrical survey made available by municipal offices (Figure 6-7), allowed to evaluate the most critical areas where to fix the sensors for environmental measures.

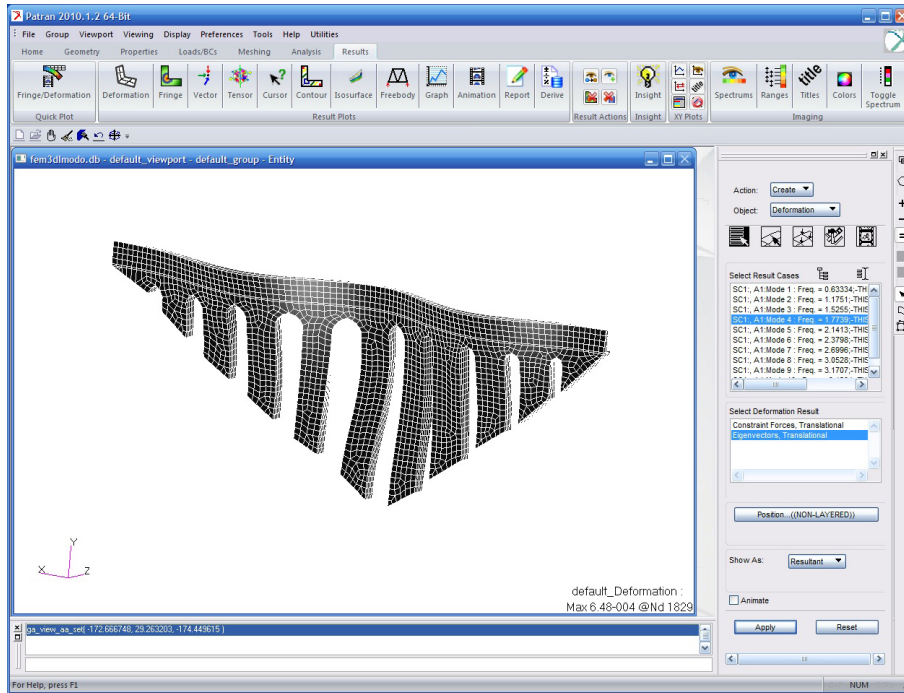


Fig. 5
Preliminary 3D FE
model of the Bridge
of the Towers
in Spoleto

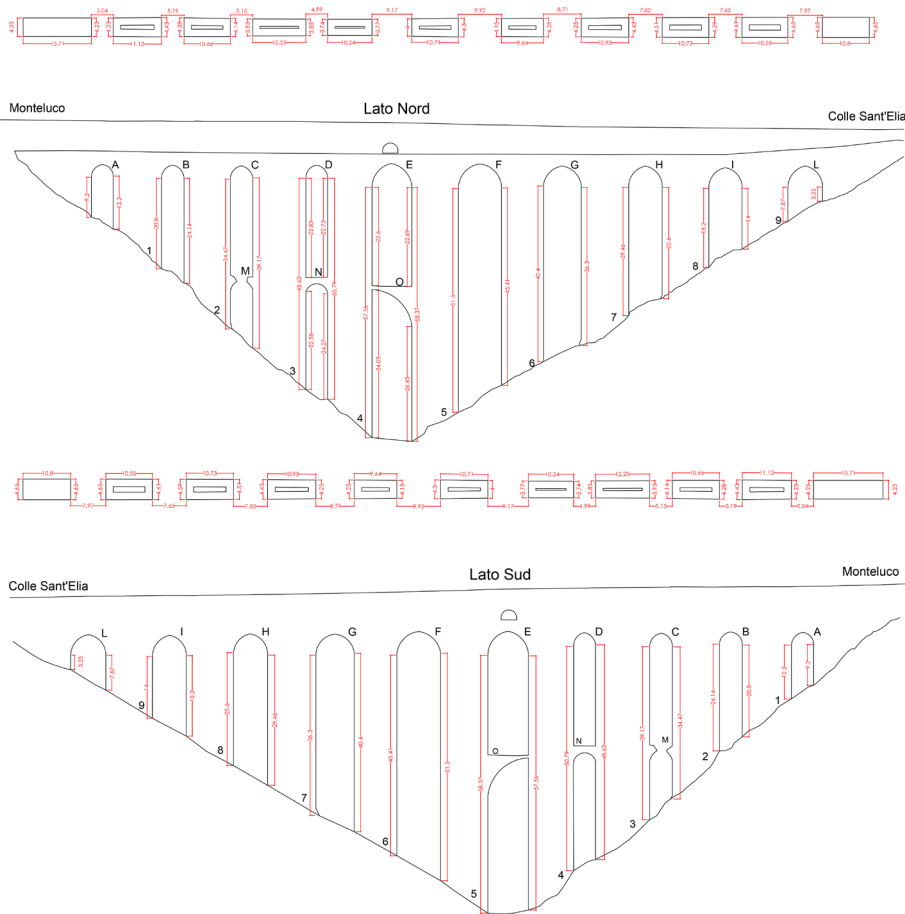


Fig. 6
Geometrical survey
of the North
side of the Bridge
of the Towers in
Spoleto

Fig. 7
Geometrical survey
of the South
side of the Bridge
of the Towers in
Spoleto

The preliminary FE model was integrated by additional geometric details at the basis of the piers, obtained by 3D photogrammetric reconstruction (Figure 8). The numerical results were then compared in terms of frequencies and modal shapes with experimental ambient vibration data, which were processed by Operational Modal Analysis (OMA) (Mongelli, M., 2017).

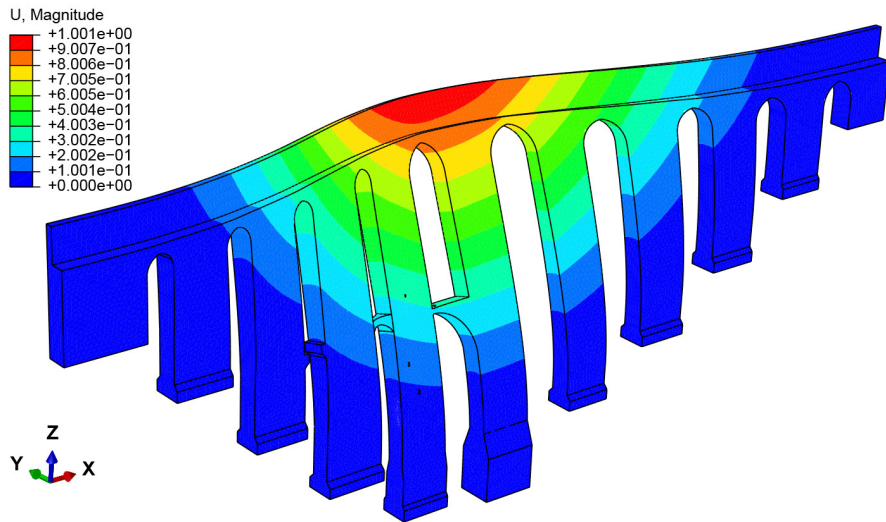
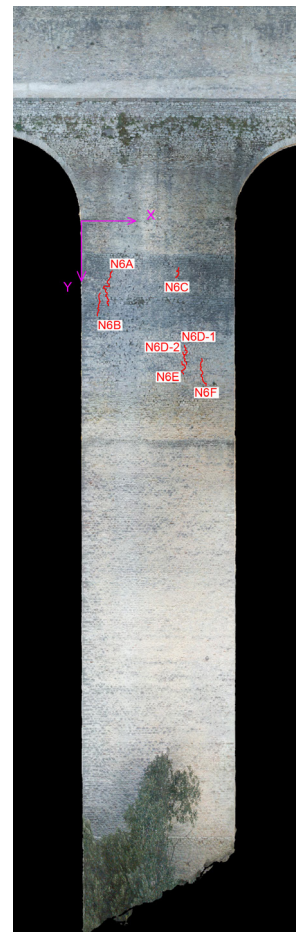


Fig. 8
FE Modal Analysis
with the basis of
the piers

Fig. 9
Crack pattern on
the pier number 6



Each identified crack was taxonomically defined in a unique and easy way (Figure 9). The indication of the damage class was defined in accordance with the macro seismic European EMS98 (http://media.gfz-potsdam.de/gfz/sec26/resources/documents/PDF/EMS-98_Original_english.pdf), also used by the 2nd GNDT level cards, considering five levels of damage (low, moderate, high, very high and collapse).

Conclusions

Finally, it is important to highlight that by IT@CHA Virtual Lab all the users will be able to develop numerical models and manage 3D data to support the preservation phase of cultural assets, taking advantage of some hardware and software capabilities for monitoring and planning diagnostic interventions. Moreover, the Virtual Lab makes available tools useful in the entire numerical modelling chain, from Digital twins to FE analysis, according to FAIR principles (for the integrated use of devices), focused on interoperability, scalability, and versatility. In the last analysis, it has been shown that the sharing of computing services and GARR network resources, using ENEA-GRID CRESCO infrastructure, is absolutely in accordance with the policies for green deal, and supports digital transition and the technological transfer.

References

Mongelli, M. et al., (2017), 3D Photogrammetric reconstruction by drone scanning for FE analysis and crack pattern mapping of the “Bridge of the Towers”, Spoleto, in Proc. of 5th International Conference on Mechanics of Masonry Structures Strengthened With Composite Materials (MURICO5), Bologna 28-30 June, Italy, published in Key Engineering Materials, Vol. 747 ISBN: 978-3-0357-1164-6, DOI: 10.4028/www.scientific.net/KEM.747.423.

Authors



Marialuisa Mongelli marialuisa.mongelli@enea.it

Chemical Engineering, PhD in Metallurgy. Researcher. She currently works in the team of HPC Laboratory of ENEA. Her activity is mainly directed in the cultural heritage field especially in the definition of virtual platforms for remote access to instrumentation for research services, consulting and qualification, as well as scientific computing provided by ENEA large plants and / or equipment. She is expert in numerical models' definition, from Digital Twin to FEM, and data analysis integration.

Beatrice Calosso beatrice.calosso@enea.it

PhD in Architecture History and Restoration. Degree in Arts and Humanities. Since 2003 employed in the ICT lab of the Italian National Agency ENEA. Expert in scientific communication and technological transfer. Communication specialist in funded projects on the use of technologies for CH diagnostic and conservation. Content Web Manager of project websites. Content editor and tutor in blended-learning professional training, about CH digital cataloguing instruments.





Samuele Pierattini samuele.pierattini@enea.it

High school graduation in Computer Science. Since 1989 employed at ENEA as a computer expert in applications and systems for advanced 3D visualization and 3D acquisition. Development of remote 3D visualization systems in GRID/CLOUD for the realization of ICT platforms and virtual labs. Participates in the PNRA –Italian National Project of Research in Antarctica in the activities of air photogrammetric surveys and aerial photographic images for environmental monitoring.

Marco Puccini marcopuccini@pm.me

Graduated in Physics from the University of Rome "La Sapienza", worked with CNR INM, DIMA and the Department of Molecular Medicine of "La Sapienza" as a data scientist. As a Research Fellow at the ICT Division of the TERIN Department of ENEA he was mainly involved in data management, development of cloud software solutions in different projects such as Mission Innovation IEMAP, DTECH, ECODIGIT, RAFAEL.



Giovanni Ponti giovanni.ponti@enea.it

Director of ENEA's ICT division and member of GARR CTS. PhD in Computer Engineering in 2010 (University of Calabria) focused on data modelling and data mining algorithms. ENEA researcher for 12 years with experiences in HPC, Data Science, Cloud Computing, Big Data and Data Analytics. Author of prestigious scientific publications. Member of the program committee of conferences and reviewer for prestigious journals and conferences.

IIT Dataverse: il repository dell'Istituto Italiano di Tecnologia per la conservazione e la condivisione dei dati FAIR della ricerca

Valentina Pasquale, Alessandro Bruchi, Ugo Moschini, Elisa Molinari, Daniele Rossetto, Francesca Cagnoni, Stefano Bencetti
Istituto Italiano di Tecnologia

Abstract. Nel riconoscimento dell'importanza di una gestione corretta e responsabile dei dati di ricerca, l'Istituto Italiano di Tecnologia ha implementato un data repository che consente ai ricercatori di documentare, conservare e condividere i dati che supportano le scoperte scientifiche dell'Istituto. Il repository è basato sul sistema open source Dataverse, adottato da numerose istituzioni a livello mondiale, che supporta i principi FAIR ed abilita il riuso dei dati per nuove ricerche. In questo articolo presentiamo il processo che ha portato all'adozione di questo strumento, la sua configurazione ed infrastruttura tecnica, i termini e le linee guida di utilizzo, ed infine i primi risultati e i possibili sviluppi futuri

Keywords. repository, Dataverse, research data management, FAIR data, storage

Introduzione

Negli ultimi anni, l'Istituto Italiano di Tecnologia ha dedicato risorse al consolidamento delle attività per la gestione dei dati della ricerca (Research Data Management), che comprende le varie fasi del ciclo di vita dei dati, dalla pianificazione alla conservazione, alla condivisione e al riuso. Tale supporto è una collaborazione delle Direzioni Organizzazione della Ricerca e ICT, con il contributo della Direzione Affari Legali.

In questo ambito, è stato implementato un repository per i dati della ricerca "definitivi", ovvero dati che sono selezionati, curati e pronti per la conservazione a lungo termine e/o condivisi in modo aperto in una forma definitiva o pseudo-definitiva. Lo strumento che è stato giudicato più idoneo per realizzare tale repository è il sistema Dataverse, un'applicazione web open source sviluppata da Harvard University nell'ambito delle Scienze Sociali e nata allo scopo di condividere, conservare, citare, esplorare ed analizzare i dati della ricerca. Questo sistema è facilmente configurabile e integrabile con altri sistemi informativi, supporta i principi FAIR (Wilkinson et al., 2016) per la gestione responsabile dei dati della ricerca ed è adottato da numerose Istituzioni a livello mondiale.

1. Identificazione dei requisiti e progetto pilota

A luglio 2019 è stato proposto un sondaggio, rivolto a ricercatori e tecnici di IIT (1547 destinatari), allo scopo di mappare le pratiche di Research Data Management (RDM) e il livello di consapevolezza in questo ambito. Il sondaggio era mirato all'individuazione

dei principali punti critici e alla scelta di uno strumento che rispondesse alle esigenze dei ricercatori in tema RDM. Il sondaggio è stato basato in larga parte su quello condotto nel 2017 e 2018 presso TU Delft, EPFL, University of Cambridge e University of Illinois, con alcune modifiche e aggiunte per tenere conto delle specificità istituzionali (TU Delft Data Stewards et al., 2018). Sono state raccolte 557 risposte, dalle quali è emersa l'esigenza di dotarsi di un repository istituzionale per supportare la conservazione dei dati e la condivisione interna ed esterna con i collaboratori. Pertanto, abbiamo cercato di individuare uno strumento flessibile, open source e che offrisse una Application Programming Interface (API) sufficientemente sviluppata per configurare il sistema programmaticamente in modo da catalogare i dati sulla base dell'organizzazione dell'Istituto e definire uno schema di autorizzazioni e permessi sufficiente ad abilitare la condivisione selettiva. Lo strumento scelto, Dataverse, offre tutte queste funzionalità ed inoltre abilita la condivisione aperta dei dati in modo aderente ai principi FAIR, uno dei principali requisiti posto dagli enti finanziatori (es., la Commissione Europea). Pertanto, nell'aprile 2020 è stato avviato un progetto pilota di 8 mesi che ha coinvolto 8 Linee di Ricerca, 2 per ciascuno dei Domini nei quali si articola l'attività dell'Istituto: Computational Sciences, Life Technologies, Nanomaterials, e Robotics. Lo scopo del progetto pilota era testare lo strumento Dataverse proponendolo ai ricercatori, i quali sono stati intervistati successivamente alla fine del pilota per raccogliere suggerimenti sulla configurazione e su possibili miglioramenti.

2. Configurazione dell'infrastruttura del sistema di produzione

Il repository IIT Dataverse è stato lanciato in produzione nel maggio 2021. Durante il primo semestre dell'anno abbiamo lavorato alla messa a punto del sistema, sia dal punto di vista tecnico, lavorando sull'infrastruttura hardware e sulla configurazione della parte applicativa, sia da quello normativo, redigendo i termini di utilizzo (<https://short.iit.it/terms>) e le linee guida per la creazione dei dataset (<https://doi.org/10.48557/EFFP8S>). Dataverse è stato installato su un'infrastruttura dall'architettura scalabile (Figura 1), basata su 3 virtual machines (VMware vCenter), per un totale di 12 CPU, 40 GB RAM, 200 GB HD e 10 TB di storage dati. Il sistema di produzione è replicato su un'istanza di staging, non accessibile dall'esterno e dotata di uno storage dati più limitato, che viene utilizzata per sviluppo e testing. Il sistema è attualmente aggiornato alla versione 5.10.1.

Da un punto di vista applicativo, IIT Dataverse è integrato con il sistema di autenticazione single sign-on dell'Istituto. Tramite API è stata automatizzata la creazione giornaliera dei dataverse associati alle Linee di Ricerca, dell'aggiunta degli utenti e dei loro permessi. IIT Dataverse è integrato con OpenAIRE, al quale invia i metadati necessari all'identificazione dei dataset e alla loro associazione con gli ORCID degli autori e i grant ID. Inoltre, è indicizzato in re3data.org e il nostro Istituto fa parte del Global Dataverse Community Consortium.

3. Linee guida e supporto

Dal punto di vista della formazione, è stata creata una pagina dedicata nella intranet di IIT che mette a disposizione materiale (presentazioni, linee guida stampabili, FAQ, video

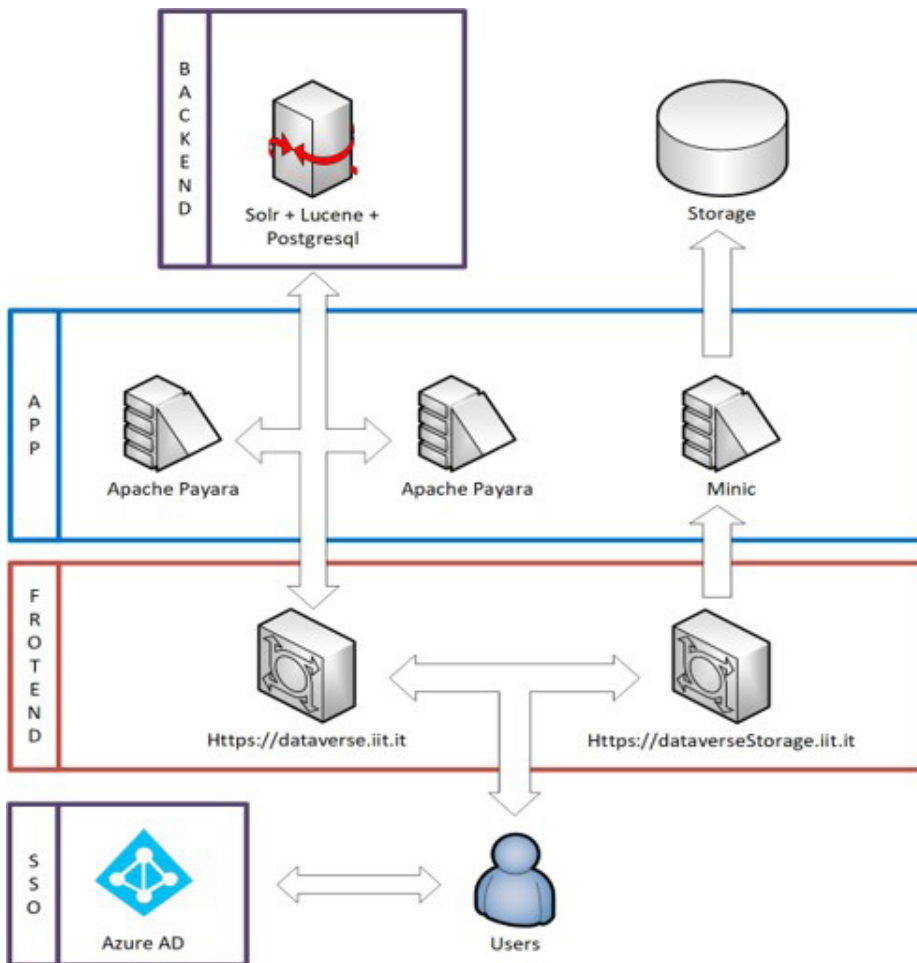


Fig. 1
Infrastruttura tecnica del repository IIT Dataverse.

dimostrativi, etc.) per assistere i ricercatori nel deposito dei dati secondo i principi FAIR. Nei mesi immediatamente successivi al lancio è stata abilitata la pubblicazione dei dataset sotto la supervisione e con il supporto del team RDM. È stata messa a punto una procedura di data curation, la quale prevede che i ricercatori attivino una “richiesta di pubblicazione dataset” attraverso un form e sottomettano il dataset per revisione. La richiesta viene presa in carico da un data steward che revisiona il dataset prima della pubblicazione sulla base di una checklist, in parte ispirata alla CURATED checklist della Data Curation Network e alla Curator Guide del repository DataverseNO. Eventuali richieste di aggiornamento o modifica del dataset vengono riportate al ricercatore, il quale si fa carico di implementarle al fine della pubblicazione. I termini di utilizzo del sistema sono conformi alle policy e procedure di IIT, in particolare 1) per la sicurezza delle informazioni (cyber-security), che definisce e regola il trattamento dei dati sulla base della classe di rischio informatico associata, e 2) per la gestione dei dati personali secondo le leggi vigenti. Per scelta IIT Dataverse non è un repository per la condivisione dei testi delle pubblicazioni, che sono esplicitamente esclusi dalla tipologia di prodotto che è possibile caricare e condividere in questo sistema.

4. Conclusioni

Allo stato attuale (settembre 2023), IIT Dataverse contiene 30 dataset di ricerca, di cui 16 unpublished e 14 pubblici (escludendo dataset creati dal supporto RDM). Si rileva che, nonostante le periodiche campagne di formazione e sensibilizzazione interna, il numero di dataset pubblici rimane limitato in ordine di grandezza rispetto alla produzione di pubblicazioni scientifiche su journal dell'Istituto (> 1000/anno). È necessario considerare che IIT Dataverse non è l'unico strumento usato dai ricercatori di IIT per la condivisione dei dataset di ricerca, che sono pubblicati anche tramite Zenodo, Figshare e altri repository disciplinari (es., eBrains) già a disposizione della comunità scientifica. IIT Dataverse non viene presentato ai nostri ricercatori come l'unica soluzione disponibile per la condivisione aperta. Tuttavia, ci si pone come obiettivo di promuoverne ed incrementarne l'utilizzo, principalmente per la conservazione a lungo termine e il riutilizzo del patrimonio di dati di ricerca dell'Istituto, migliorando il supporto alla compilazione dei metadati e rendendolo uno strumento sempre più integrato all'interno dei flussi di lavoro.

Inoltre, IIT si è recentemente dotato di una nuova infrastruttura di storage centralizzata scale-out che ha l'obiettivo di aumentare la capacità di gestione dei dati, un loro uso sempre più efficace e che apre alla possibilità di applicare modelli di artificial intelligence ai dati stessi. IIT Dataverse sarà quindi integrato nel ciclo di vita del dato e verrà dotato di maggiore capacità, per garantire una migliore gestione di dataset di grandi dimensioni, e di termini di utilizzo aggiornati che abiliteranno anche la condivisione in forma restricted.

Riferimenti bibliografici

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

TU Delft Data Stewards, EPFL Library Research Data Team, Krause J., Lambeng N., Andrews H., Boehmer J., Cruz M., van Dijck J., den Heijer K., van der Kruijk M., & Teperek M. (2018), Quantitative assessment of research data management practice (Version 3) [Data set], Zenodo. <https://doi.org/10.5281/zenodo.3380332>

Autori



Valentina Pasquale valentina.pasquale@iit.it

Valentina Pasquale è Research Data Management Specialist all'Istituto Italiano di Tecnologia, dove si occupa dei servizi di supporto per la gestione dei dati della ricerca e la scienza aperta. Dal 2019 al 2021 è stata co-chair della rete di Data Stewardship Competence Centers in GO FAIR. Prima di occuparsi di dati della ricerca, ha ottenuto un dottorato in "Humanoid Technologies" all'Università di Genova e ha lavorato per diversi anni nel campo delle Neuroscienze Computazionali.

Alessandro Bruchi alessandro.bruchi@iit.it

Alessandro Bruchi è laureato in Ingegneria Elettronica e specializzato in progettazione micro-elettronica presso l'Università di Padova. Ha lavorato come consulente presso Engineering SPA

su progetti bancari e assicurativi, acquisendo esperienza in sistemi, reti e Linux. Dal 2007 lavora all'Istituto Italiano di Tecnologia e dal 2015 è responsabile dell'ufficio Cybersecurity, competente per la protezione delle infrastrutture e dei dati.



Ugo Moschini ugo.moschini@iit.it

Ugo Moschini, dopo aver lavorato all'Agenzia Spaziale Europea su algoritmi di compressione dati, ha conseguito un Dottorato in Informatica dall'Università di Groningen in analisi e classificazione di dataset di astronomia. Dal 2017 lavora nel Data Analysis Office dell'Istituto Italiano di Tecnologia su attività di analisi, visualizzazione e modellazione di dati riguardanti la produzione scientifica dell'istituto, con attenzione anche alle pratiche di Research Data Management e Open Science.

Elisa Molinari elisa.molinari@iit.it

Elisa Molinari ha conseguito un dottorato in Bioingegneria e Bioelettronica al DIST di Genova, progettando e sviluppando una piattaforma hardware e software per la gestione di dati neuroscientifici. Ha poi ottenuto un postdoc all'IIT e collaborato con università e l'ospedale pediatrico Gaslini nel campo dell'analisi di immagini biomedicali. È adesso Lead Data Analyst all'IIT di Genova e si occupa del coordinamento e dello sviluppo di sistemi di monitoraggio dell'attività di ricerca.



Daniele Rossetto daniele.rossetto@iit.it

Daniele Rossetto Casel si è laureato in Informatica presso l'Università di Genova nel 2004 e dal 2005 al 2009 ha lavorato presso lo stesso Ateneo come amministratore di rete e di sistema. Dal 2010 lavora presso l'Istituto Italiano di Tecnologia e dal 2014 ricopre il ruolo di responsabile dell'ufficio Gestione Infrastrutture, supportando l'Istituto e gli utenti su tutte le tematiche di infrastruttura IT.



Francesca Cagnoni francesca.cagnoni@iit.it

Francesca Cagnoni ha conseguito un dottorato in Immunologia Clinica e Sperimentale ed una specializzazione in Allergologia e Immunologia Clinica all'Università di Genova. Successivamente ha lavorato in un'azienda biotech prima di passare al supporto alle attività scientifiche. Oggi è Direttore del Research Organization Directorate di IIT, dove coordina un team trasversale che comprende gli uffici di Data Analysis, Projects, Research Agreement, e Outreach and Digital Production.



Stefano Bencetti stefano.bencetti@iit.it

Stefano Bencetti si laurea in Ingegneria Elettronica presso l'Università degli Studi di Genova e inizia la sua esperienza lavorativa nel settore ICT nel 1997 lavorando presso il Dipartimento di Matematica e il Dipartimento di Scienze dell'Informazione dell'Ateneo genovese. Dal 2009 è Direttore ICT presso l'Istituto Italiano di Tecnologia.



FOSSR-Fostering Open Science in Social Science Research: Building Italy's Innovative Open Cloud Infrastructure

Serena Fabrizio¹, Mario Sicuranza², Mario Ciampi², Emanuela Reale¹

¹Research Institute for Sustainable Economic Growth of the National Research Council of Italy, ²Institute for High Performance Computing and Networking of the National Research Council of Italy

Abstract. FOSSR project is dedicated to creating an Italian Open Science Cloud, inspired by the European Open Science Cloud (EOSC). A central strategy of FOSSR revolves around leveraging Research Infrastructures (RIs) in the social sciences to ensure the availability of data adhering to the FAIR principles. FOSSR aligns with the broader mission of advancing Open Science in Italy, aiming to provide a framework of tools and services to support the social science academic community. This includes the integration of the Italian nodes of three existing RIs: CESSDA-ERIC, SHARE-ERIC, and RISIS, coordinated by the National Research Council of Italy (CNR). The FOSSR project strategically integrates cloud computing infrastructure and a Network of Data Centers, primarily situated in southern Italy which strengthens data processing capabilities. This network will enhance local computing capabilities and support the advanced data analytics needs of social science researchers. The federated platform offers various interfaces, and a Virtual Research Environment (VRE), promoting open, collaborative, and streamlined data access.

Keywords. Italian Open Science Cloud, Research Infrastructures, Cloud Computing, Social Science Research, Innovation

Introduzione

FOSSR - Fostering Open Science in Social Science Research aims to create an Italian Open Science Cloud, inspired by EOSC (European Open Science Cloud), in which to integrate cutting-edge tools and services dedicated to exploring issues pertinent to the evolution of contemporary societies' economic and social landscape. A key strategy to achieve this objective revolves around the exploitation of research infrastructures (RIs) in the social sciences, which facilitate the availability of data that adhere to FAIR (Findable, Accessible, Interoperable, Reusable) principles.

FOSSR embraces the overarching theme of advancing Open Science in Italy with the aim of establishing a framework of tools and services for the social science academic community. This involves the Italian nodes of three RIs in social sciences coordinated by CNR, including CESSDA-ERIC (Consortium of European Social Science Data Archives), SHARE-ERIC (Survey of Health, Ageing and Retirement in Europe), and RISIS (Research Infrastructure for Science and Innovation policy Studies), as well as ISTAT (Italian Institute for Statistics).

FOSSR primary objective is to promote widespread knowledge and awareness of data and methods used in empirical social science among multiple audiences. This is accomplished by providing (i) systematic and organized knowledge about available social science data resources in Italian data archives, particularly CESSDA Archive, which was the subject of a major infrastructural proposal; (ii) resources to support methodological advancements in data collection and analysis, which are particularly important for RISIS to understand the design, implementation, and outcomes of research and innovation policies, thereby improving the robustness of empirical evidence produced for policymakers and addressing new research questions; and (iii) tools and services to make advanced probability panels for longitudinal analyses publicly available to support important surveys such as SHARE, complemented by a network of online laboratories. The integration of these resources will directly contribute to the realization of Open Science for social science scholars, accompanied by a significant scientific training program for the production and analysis of social science based on FAIR empirical data.

1. FOSSR strategic nodes

Building a thematic network of existing RIs is a unique opportunity to improve their quality with high innovative services and resources, not existing in Italy, which can have a special value to support research and innovation in line with the main objectives of the PNRR – Action M4C.2.

The project will integrate hardware and software tools together with methodologies aligned with research approaches rooted in e-science, behavioral economics, and computational social sciences. From a practical point of view, it foresees significant progress in features such as data collection, integration, curation and sharing, the establishment of a survey facilitator, the development of a social listening framework, and the launching of an artificial population simulation facility.

In the operational context, this framework aims to build a unified knowledge-sharing platform, serving as a centralized gateway for all the tools and services provided by the Italian nodes of social science infrastructures. To achieve this objective, a nationwide platform will be conceptualized and constructed, leveraging a distributed cloud computing infrastructure. This platform will facilitate the establishment of a unified and cohesive system, encompassing national infrastructure nodes (such as CESSDA, RISIS, and SHARE), in addition to internally generated data and statistical information from ISTAT.

The mentioned RIs are also important for non-academic users, such as stakeholders and policymakers. Despite the number of users coming from Italy is relevant, there is a need to improve the services and data provided to address new emerging research questions related to the changes of the economy and society, as well as to the transformative turn of science and innovation.

The integration of this pool of resources will concretely contribute to the realization of Open Science for social science scholars, accompanied by a major scientific training program on social science research methods and tools based on FAIR empirical data.

The main innovation for all the RIs involved in the thematic network are:

- Promoting the use of data for the social sciences among scholars and non-academic users and assisting data users to exploit data and resources.
- Providing innovative tools to collect, explore, analyze, and harmonize data.
- Building a distributed set of data centers to store and distribute data and resources derived from the combined activities of the thematic network.
- Training new generations of researchers/data users and disseminate the results deriving from projects developed by using FOSSR resources, developing Master courses and funding PhD positions for research on economic and societal change (in particular in the South of Italy), which allow the creation of a new pool of scholars in Data Science.

FOSSR aims to produce in this way a strong impact on the Italian communities located in the South, also through an important investment in large data centers that put the Open Cloud in this part of the country. This way, it might become a number of infrastructural resources at disposal of other public research organizations, Universities and possibly private operators and firms.

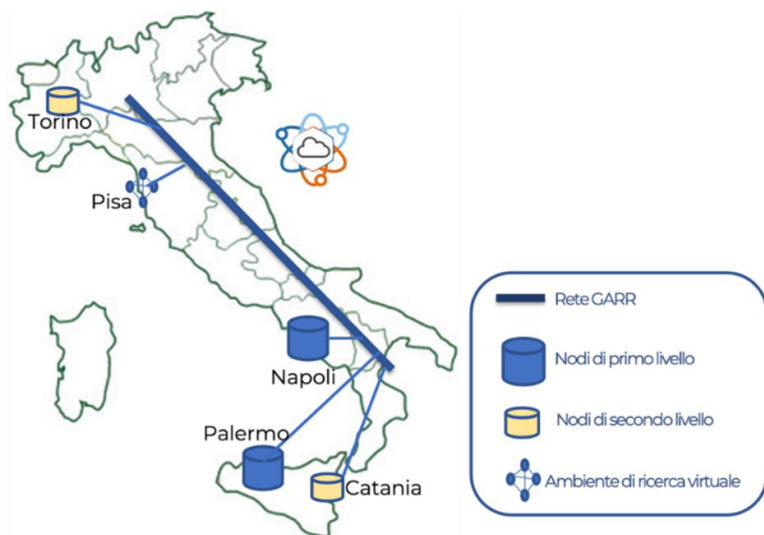
2. The heart of FOSSR: Cloud Computing Infrastructure and Network of Data Centers

The Open Cloud is developed on the top of a network of distributed data centers located in 4 strategic areas including 2 medium-large nodes (Naples and Palermo) and 2 small nodes (Turin, and Catania).

2.1 Cloud Computing Infrastructure

Through the cloud computing infrastructure, researchers will be empowered with a user-friendly, consolidated entry point that grants them access to computational services. It will also provide the means to archive and reach virtual laboratories, enabling the utilisation, processing, and public dissemination of datasets and outcomes. This unified platform will enable users to effortlessly access all the available resources and tools via a solitary entry point, employing a Single Sign-On (SSO) identification mechanism, by offering a single access point and a series of innovative services for the collection, mana-

Fig. 1
Network of
Data Centers



gement, and analysis of economic and social data in compliance with the FAIR principles (Wilkinson, 2016).

As shown in Fig. 1, the platform will provide an avenue to access data in an open, collaborative, and streamlined manner facilitated by shared interfaces.

These interfaces will accommodate various tiers of access, specifically marketplace, programmatic interfaces, query language, as well as a Virtual Research Environment (VRE):

- **Marketplace.** Developed within a multilingual web portal, the FOSSR marketplace will afford social science researchers the capacity to access and distribute applications, services, informative resources, and datasets: it allows functionalities for searching, storing, retrieving, and processing the content through a well-defined catalogue. Users will be able to publish, analyse, and explore content via a comprehensive catalogue.
- **Programmatic Interfaces:** The information encompassing data, metadata, and service catalogue can be interacted with by external applications through machine-to-machine communication protocols rooted in OpenAPI/Swagger REST APIs.
- **Query Language:** The cloud infrastructure will facilitate the querying of both data and metadata through the utilization of standardized languages and protocols. As a foundational language and protocol for this purpose, SPARQL will be employed, enabling the querying of data and metadata over HTTP(s).
- **Virtual Research Environment (VRE):** VRE will serve as a catalyst for collaborative research and the promotion of open science principles, employing an array of datasets and analytical tools, supplemented by corresponding services. The FOSSR VRE will be constructed upon the foundation of the established electronic infrastructure known as D4Science, a venture overseen by CNR-ISTI.

The upcoming Cloud Platform will serve as a valuable tool for unifying servers and hardware resources to construct a cohesive system. This federated infrastructure will be established upon the existing hardware infrastructure, specifically the data center network. This network enhancement will be carried out through initiatives aimed at upgrading the four data centers.

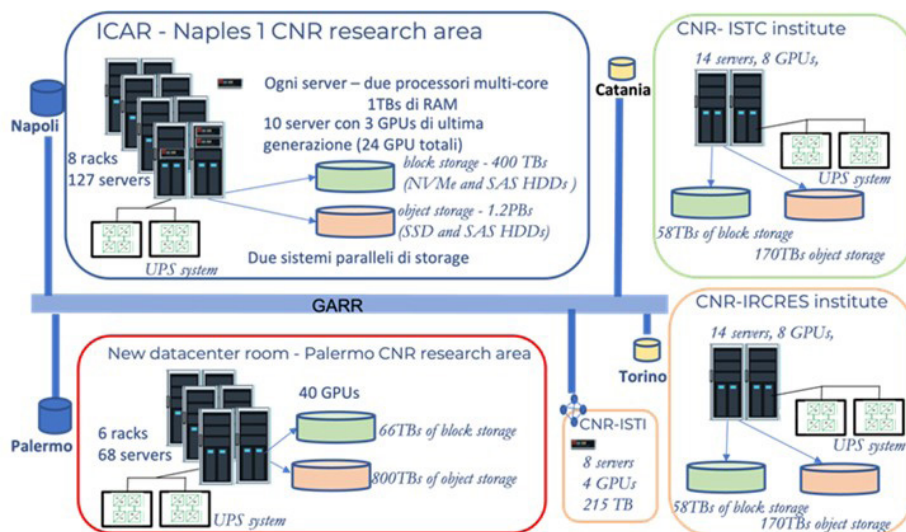
2.2 Network of Data Centers

The network of data centers includes 2 medium-large nodes (CNR-ICAR, Naples and CNR-ICAR, Palermo) and 2 small nodes (CNR-IRCrES, Turin and CNR-ISTC, Catania). The former is referred to as First Level Nodes (1LNs), while the latter as Second Level Nodes (2LNs). Additional hardware will be purchased in Pisa to deliver a Virtual Research Environment (VRE) on a dedicated IT infrastructure. The two 1LNs will be designed to provide access to hardware and software resources for high-performance computing and big data storage by means of cloud computing technologies.

The distinction between first and second level mainly regards the size and number of machines, where the application services made available to the infrastructure are similar. The main difference is that the first-level nodes will deal more with services for external exposure (web portal, marketplace, access point and security), but all nodes will have the same levels of security and Service Level Agreements (SLAs).

The data center at the Naples site of CNR-ICAR will be the largest among the first-tier data centers and will have to feature hardware components capable of managing the cloud system, guaranteeing its availability and functionality while at the same time providing a computing infrastructure adequate for the operation of services and advanced tools for data analysis using artificial intelligence algorithms, in particular based on Deep Learning. The Naples 1LN will be hosted in the ICAR data center located in the Naples 1 CNR research area. The data center will host 8 racks with 100 servers. Each server will be equipped with two multi-core processors and 1TBs of RAM allowing different kinds of computing workflows: data-intensive, CPU-intensive, parallel computing and virtualization. Some of these servers will host the latest generation Graphics Processing Units (GPUs) to run applications to handle massive AI and DL workloads. Each of these servers will host 4 units to allow their use in parallel for training deep neural networks and shorten the time required for data-intensive applications. The computing server cluster will be connected to two separate parallel storage systems: one dedicated to block storage with 400 TBs composed of nvme and SAS HDDs and a second one dedicated to object storage with 1.2PBs with SSD and SAS HDDs. The servers and the storage systems will be interconnected with a high throughput and low latency network technology specialized for HPC systems. The presence of two 1LNs will provide the cloud infrastructure with enough redundancy to allow high availability of its cloud functionalities and the possibility to replicate data across the sites for disaster recovery; it will also enable the development of scalable applications, capable of handling high loads and traffic spikes, with multi-region architecture for resilience to local outages. The complete configuration is schematized in Fig. 2.

Fig. 2
Data Centers
Configuration



2.3 GARR Connection and Management Network

As mentioned, the different nodes will be directly connected by the GARR network, each node within it being configured with different networks.

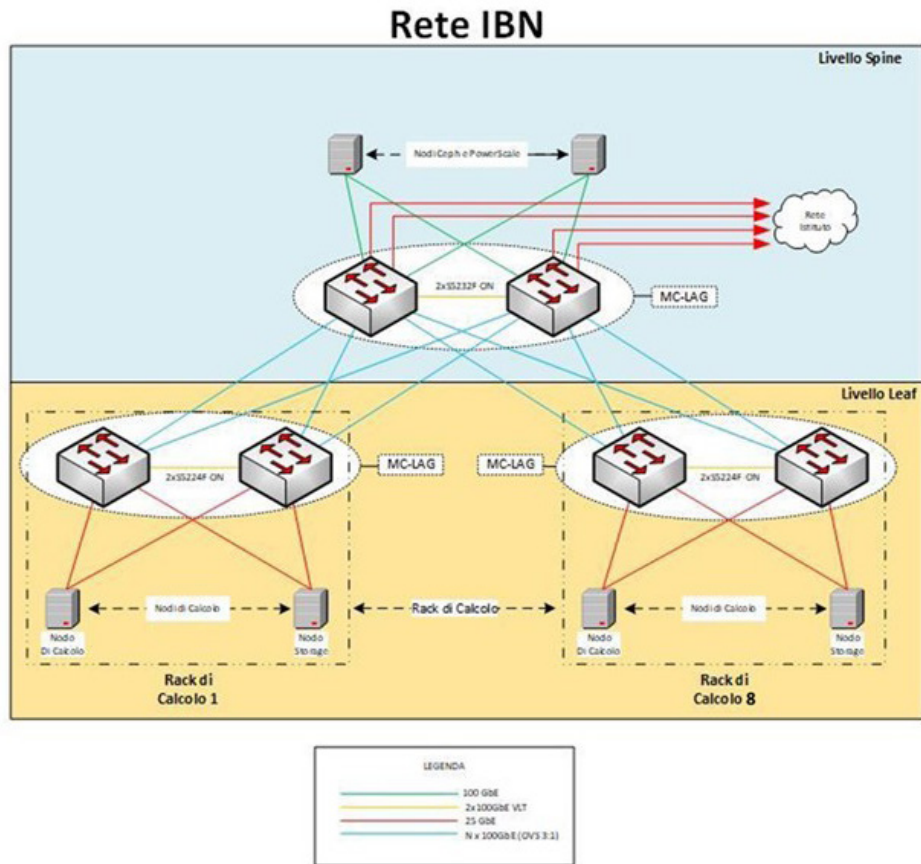
As shown in Fig. 3, the In-Band Management Network (IBN) allows high-speed intercon-

nection of the storage nodes (of the Ceph servers) of the Cloud partition as well as data traffic for communications useful for infrastructure management. This network is configured in a fully redundant spine-leaf topology

The OOBN interconnects the Baseboard Management Controllers (BMCs) of all nodes and the out-of-band (OOB) management interfaces of all other infrastructure components (storage systems, network equipment, PDUs, etc.). This network is configured in a spine-leaf topology and is redundant at the spine level.

The LLN interconnects all the computing nodes of the HPC partition of the server cluster. This network is realized with RockPort technology.

Fig. 3
In-Band
Management
Network



3. Conclusions

The FOSSR project represents a significant step forward in the field of Open Science in Italy and will help to foster the development of social sciences in a fair, transparent, and accessible way. It will also help to address some of the challenges that social science researchers face when it comes to data sharing and open-access publishing.

The cloud platform for social sciences has the aim to provide innovative tools and services for the analysis of social and economic data from different systems. The utilisation of cloud computing infrastructure will facilitate the seamless integration and provisioning

of services geared towards storing, analysing, querying, retrieving, and sharing extensive arrays of social science data, all aligned with a standardised data model. The mix of these tools and services with open science practices will help to increase the visibility and impact of social science research in Italy.

References

D4Science Infrastructure, Infrastructure for the management of scientific data, <https://www.d4science.org/about-us>

Garr Network, <https://www.garr.it/it/infrastrutture/rete-nazionale/infrastruttura-di-rete-nazionale>

Reale, Emanuela, Ciampi, Mario, Paolucci, Mario, Zinilli, Antonio, Nuzzolese, Andrea Giovanni, Cerulli, Giovanni, Spinello, Andrea Orazio, De Gregorio, Daniela, Saccone, Massimiliano, & Rosati, Maria Elisa. (2022, December 21). FOSSR Kick-off meeting poster and presentations. Zenodo. <https://doi.org/10.5281/zenodo.7458327>

Rockport Networks, <https://rockportnetworks.com/>

Stilo, Maria Alessandra, Fabrizio, Serena, & Fava, Alessia. (2023). FOSSR: Driving Fair and Open Social Science Research (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.8005487>

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

Authors



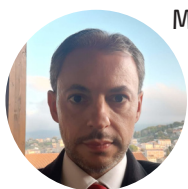
Serena Fabrizio serena.fabrizio@ircres.cnr.it

Serena Fabrizio PhD in Communication Science at Sapienza University of Rome is Researcher at CNR-IRCrES. Since 2015 with a research fellowship at CNR-IRCrES worked on several European and national research projects. In particular in the management of Training and Communication activities of RISIS project, and in the research team of a PRIN coordinated by CNR-IRCrES. Main research interests concern policies for higher education and governance systems, analysis of public funding on R&D, evaluation of impact of Social Science and Humanities research. She is involved in the FOSSR project, on Community building and training activities.

Mario Sicuranza mario.sicuranza@icar.cnr.it

Mario Sicuranza, PhD, MSc, is a Research Technologist of the Institute ICAR CNR. He received a Ph.D. degree in Information Engineering on Cybersecurity for Health Information System in 2016. His research interests include e-health, web services, and security architectures. He collaborates with the Italian Presidency of the Council of Ministers, the Ministry of Health, and the Ministry of Economy and Finance, on the National Interoperability Framework for the interoperability of the Electronic Health Record regional systems. He is co-author of more than 50 scientific papers. He leads the WP7 on developing the data center network in FOSSR.





Mario Ciampi mario.ciampi@icar.cnr.it

Mario Ciampi is a Senior Technologist at CNR-ICAR, where he works on technology research in the field of information systems, by leading the "System Interoperability and Management" technology group. He has held numerous leadership roles in research projects co-funded by European Commission, Italian Ministries, and Regions. He acts as an evaluator of research projects for the European Union, international advisory boards, and national authorities. He leads the WP6 on Cloud Computing in FOSSR.

Emanuela Reale emanuela.reale@ircres.cnr.it

Emanuela Reale is Director of CNR IRCrES. Her research area is the study of the public sector research institutions and policies, with particular reference to university policy, governance, funding mechanisms, methods and tools for university and research assessment, science and technology indicators. She has worked in numerous national and international projects as a principal investigator or coordinator. She is member of the Consortium Management Committee of RISIS, of the Board of Eu-SPRI, of the Board of CHER, and of the Board of ENID. She is the Scientific Coordinator of the FOSSR project.



A Cloud Computing infrastructure for the Competence Centre for the Conservation of Cultural Heritage

A. Alkhansa¹, A. Bombini², A. Costantini¹, L. dell'Agnello¹, F. Taccetti²

¹Center for Research and Development on Information and Communication Technologies (CNAF), ²Istituto Nazionale di Fisica Nucleare (INFN)

Abstract. Within the context of the Competence Centre for the Conservation of Cultural Heritage (4CH) project, the design and deployment of a Platform-as-a-Service Cloud infrastructure as a prototype for the first European Competence Centre of Cultural Heritage has begun. On top of this, some services used and developed for the cultural heritage community have been integrated. In this contribution, we summarise the status of the work carried out within the 4CH project activities, the services currently hosted, and the different solutions adopted. We conclude by describing the next step towards the creation of the first European cloud platform (and related services) for cultural heritage

Keywords. Cloud, Strategies and policies for the development of e-infrastructures, ICT services for Research Infrastructures, Enhancement of Cultural Heritage

Competence Centre for the Conservation of Cultural Heritage (4CH)

The Competence Centre for the Conservation of Cultural Heritage (4CH) project is an initiative that was approved in January 2021 under the DT-TRANSFORMATIONS-20-2020 call of the European Community's Horizon 2020 framework programme.

Its goal is to design and prepare for an European Competence Centre (CC) for the Conservation of Cultural Heritage which will work proactively for the preservation and conservation of cultural heritage (CH). The Competence Centre will address several topics, such as how cultural heritage is managed, the risks it may be exposed to, the technologies that can be used for its conservation and how it shall be exploited.

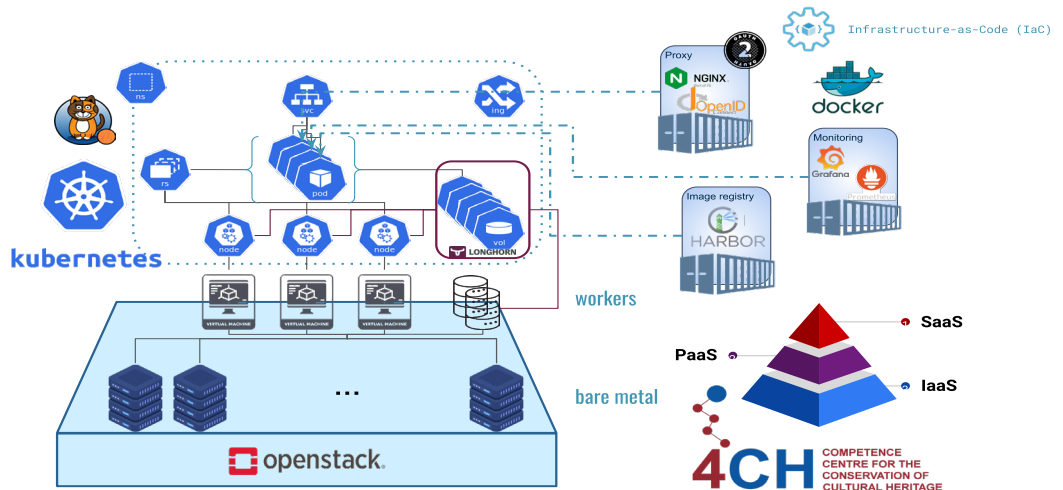
Towards the development of a 4CH Cloud Platform

During the 4CH project activities, an important effort has been spent to design a proper architecture aimed at hosting all the elements needed to provide users with the capability of using different software services, made available via a platform in order to let them be able to use resource infrastructures in a seamless and transparent way.

To host the 4CH services for cultural heritage, a cloud-based infrastructure has been implemented as a pilot infrastructure hosted on the OpenStack Cloud at CNAF, the national center of INFN (Italian Institute for Nuclear Physics) dedicated to Research and Development on Information and Communication Technologies. The adopted tools and solutions are aimed at providing distributed computing resources, network, and stora-

ge. Moreover, on top of the 4CH Platform, some ancillary services providing useful functionalities such as reverse proxy, image repository, monitoring and authorization and authentication, already deployed, have been also considered and integrated together into the 4CH Platform-as-a-Service.

Fig. 1
4CH platform
design and
components



The 4CH platform development adopted the so-called Infrastructure-as-Code (IaC) approach . Following the IaC approach, Kubernetes has been adopted and implemented. The platform was deployed using the Rancher Kubernetes Engine (RKE). On top of the Kubernetes cluster, some components have been deployed to improve the platform functionalities. Network service relies on Calico, while storage service relies on Longhorn. For the Authentication / Authorization workflow, the 4CH cloud platform relies on INDIGO-IAM service. INDIGO-IAM implements (i) the OAuth2.0 standard authorization framework with the OpenID Connect (OIDC) layer, (ii) the User-Group model to manage the authorization procedure. Grafana and Prometheus have been deployed to monitor both the Kubernetes cluster and the different services running on top of it. Finally, an image repository service based on Harbor has been deployed. A graphic schema of the 4CH Cloud Platform is shown in Figure 1.

Service integration and federation

To cope with the different web-based (even commercial) applications that may arise from the different 4CH project partners and related communities, a set of principles and policies have been considered to define the requirements related to the application deployment, method and resource exposure and authorization and authentication mechanisms. In particular, an important effort has been spent to define requirements for service inte-

gration, referring to applications that can be integrated in the 4CH Platform and made available to the community via the 4CH dashboard.

On the other hand, applications that cannot be integrated in the 4CH cloud platform can be hosted outside of it, and can be made available to the community via the CHNet web portal.

By adopting this approach, Cultural Heritage applications that have specific needs (e.g. application which cannot be virtualized, in-house application with particular dependencies, commercial software etc.) can in any case be made available to the final user via the CHNet web portal adopting the authentication and authorization mechanism provided by the 4CH Platform and reach the Cultural Heritage Network communities through the Platform.

Note that a single application, comprising multiple independent services, may be deployed in a mixture of federation and integration deployments.

Next Steps

The present document describes and highlights the 4CH pilot infrastructure and its related Platform components and tools aimed at hosting and integrating the different services oriented to the Cultural Heritage community developed by the 4CH project partners. At the time being, some services provided by and for the Cultural Heritage community within the activities of the 4CH project have been ported and hosted in the 4CH Cloud Platform.

Those services and processes are part of a lively ecosystem of activities that are expected to evolve within the 4CH project to accommodate further needs and requirements that may not be raised so far.

The interaction and feedback collected from service providers and user communities will also contribute to improve the project developments as well as to define the features of the future European Competence Centre for Cultural Heritage.

Acknowledgments

The present work has been partially funded by the European Commission within the Framework Programme Horizon 2020 with the project 4CH (GA n. n.101004468 – 4CH).

References

- [1] Medici, M.; Costantini, A.; Niccolucci, F. et al. (2022). “D3.1 Design of the CH Cloud and 4CH platform” (1.0). Zenodo. <https://doi.org/10.5281/zenodo.7701438>
- [2] Bombini, A.; Alkhansa, A.; Cappelli, L.; Felicetti, A.; Giacomini, F.; Costantini, A. “A Cloud-Native Web Application for Assisted Metadata Generation and Retrieval: THESPIAN-NER”. *Applied Sciences*. 2022; 12(24):12910. <https://doi.org/10.3390/app122412910>
- [3] Bombini, A.; Castelli, L.; dell’Agnello, L.; Felicetti, A.; Giacomini, F.; Niccolucci, F.; Taccetti, F. “CHNet cloud: An EOSC-based cloud for physical technologies applied to cultural heritage”. In *Proceedings of the Conferenza GARR 2021—Sostenibile/Digita-*

le. Dati e Tecnologie per il Futuro, Online, 7–16 June 2021. <https://doi.org/10.26314/GARR-Conf21-proceedings-09>

Authors:



Ahmad Alkhansa ahmad.alkhansa@cnafe.infn.it

Research Fellow at INFN-CNAF that specializes in the orchestration of containerized workloads. His activities include designing and implementing distributed systems for hosting web applications. He is involved in the provisioning and configuration of the infrastructure dedicated to the 4CH european project, and ensuring the availability of the services deployed on top of it.

Alessandro Bombini alessandro.bombini@fi.infn.it

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 EOSCPillar (<https://www.eosc-pillar.eu/>), A-RIADNEplus (<https://ariadne-infrastructure.eu/>), and 4CH (<https://4ch-project.eu/>).



Alessandro Costantini alessandro.costantini@cnafe.infn.it

Alessandro Costantini graduated in Chemistry in 2005, in 2009 he holds a Ph.D. in Chemistry working on Clouds environments and in 2013 he holds a Ph.D. in Mathematics and Computer Science. Since 2022 he is the coordinator of the Distributed Systems group at the INFN National Centre for research in ICT technologies (CNAF) where he enriched his expertise becoming a Cloud technologies expert and participating in different Cloud-related projects. At present, he is coordinating the activity related to the Cultural Heritage Cloud Platform within the 4CH european project.



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.



Francesco Taccetti francesco.taccetti@fi.infn.it

First Technologist at INFN, expert in accelerators physics and applied physics on Cultural Heritage. He is national responsible for the INFN-CHNet, the INFN network devoted to the application of nuclear techniques to the Cultural Heritage. He is Principal Investigator for the INFN in the European projects and MIUR-financed projects, such as 4CH, ORMA, A-RIADNEplus, MACHINA, E-RIHS, and authored more than 80 papers published in international journals.



Toward Effective Cybersecurity Introduction for High-school girls

Gabriele Costa, Serenella Valiani

IMT School for Advanced Studies, Lucca

Abstract. The gender gap is one of the main concerns in the IT sector, in general, and in cybersecurity, in particular. In the last years, many initiatives were proposed targeting high-school students that may be interested in cybersecurity. Most of them rely on capture-the-flag (CTF) competitions to gradually form technical skills in an entertaining way. Despite these efforts, however, the number of girls that effectively complete these programs is still negligible. In this work, we present the results obtained by CyberTrials, a CTF-based cybersecurity program for high-school girls. The two main novelties of CyberTrials are its peculiar organization and gaming platform, which substantially differ from other similar initiatives. The preliminary results show that our methodology could effectively engage the participants

Keywords. education, hands-on training, cybersecurity, Capture the Flag, gender gap

Introduction

Gender disparity in STEM fields, driven by societal stereotypes (A. H. Master et al, 2020), a lack of role models (S. González-Pérez et al, 2020), and gender-based expectations (C. Riegle-Crumb et al 2017), is a well-documented issue (D. N. Beede et al, 2011; M. Wang et al, 2017). High school experiences contribute significantly to this gap, as studies reveal obstacles faced by young women considering STEM careers (D. Card et al, 2021; J. Legewie et al, 2014). Italy exemplifies this pattern, with 56.3% female university admissions in 2020/2021, but only 21% choosing STEM courses, perpetuating this gap post-graduation. The gender gap in cybersecurity, a STEM subset, is more pronounced, with women comprising just 25% of the workforce. This hampers innovation, necessitating diverse competencies and perspectives.

To address this, diversity initiatives must begin in high school. Mentorship programs (H. Stoeger, 2013), outreach campaigns (E. Prieto-Rodriguez et al, 2022), and tailored courses (J. M. Bystydzienski et al, 2015) are vital. Yet, cybersecurity's technical nature poses challenges for hands-on experience.

We introduce CyberTrials, an entry-level, game-based cybersecurity program for high-school girls. It introduces various research areas, allowing participants to grasp technical concepts. CyberTrials incorporates hands-on Capture The Flag (CTF) sessions alongside lectures, using gamification elements to enhance engagement (F. Fui-Hoon Nah et al, 2014).

Our results show CyberTrials effectively engages participants, imparting technical skills in Linux terminal usage, network scanning, and steganography. This approach aims to bridge the cybersecurity gender gap by inspiring the next generation of female professionals.

1. Preliminaries

Capture The Flag (CTF) competitions involve IT challenges where individuals or teams earn points for solving security-related tasks, like decrypting ciphers or bypassing authentication. Success yields a flag submitted for scoring, with the highest scorer winning. CTFs come in various formats, primarily jeopardy-style where each challenge is independent, and flags can have fixed or dynamic values. Attack-defense (AD) style has teams managing vulnerable systems, earning points through offense and defense. Quest-style CTFs, less common, integrate challenges into a narrative storyline. Solved challenges impact the game, unlocking hints or new challenges, enriching the narrative and complexity.

2. The CyberTrials Program

The CyberTrials program is designed as an entry-level cybersecurity training initiative tailored for high-school girls. The program's inception occurred approximately one year ago, and the second edition has just been concluded. Here, we do not report any details from the first edition, as it is a light-weight edition of the second one.

Enrollment in the program was open to high school students affiliated with CyberHighSchools, a nationwide network of educational institutions. The entirety of the program was conducted remotely, with the exception of the final challenge, which was hosted in Turin. Participants were not required to possess technical competencies, and notably, programming skills were not a prerequisite.

2.1 Program Structure

The CyberTrials program aimed to provide high-school girls with foundational cybersecurity skills and a comprehensive understanding of various cybersecurity aspects. The curriculum covered areas like ethics and legal considerations, computer networks, open-source intelligence, threat modeling, digital forensics, cryptography, and social engineering. Soft skills and teamwork were also emphasized.

The program began on January 30, 2023, with an introductory meeting where participants were grouped into four-member teams. The first week focused on team-building activities to help participants get acquainted and included introductory challenges to familiarize them with the platform. From the second week, weekly two-hour online lectures covered the designated topics. Expert speakers from academia and industry led each module, and recorded sessions were provided to participants.

The theoretical aspects were complemented by hands-on training, featuring a continuous CTF competition throughout the training period. Each week introduced three new challenges of varying complexity, from easy to difficult.

A unique aspect of CyberTrials was its integration of continuous storytelling, particularly evident in the second edition's quest-style CTF. The narrative centered around an anonymous web artist seeking validation of authorship for stolen and misattributed artworks. All challenges were intricately linked to this narrative, incorporating references, descriptions, hints, and additional materials. External activities, like an Instagram account associated with the fictional web artist, contributed to the narrative's development.

2.2 Program Timeline

CyberTrials began on October 10th, 2022, with online registration, closing on January 15th, 2023, with 941 enrolled students. The training phase ran from January 30th to April 15th. The CTF scoreboard froze on March 13th, ending on March 27th after Module 8. Final rankings used the CTF scoreboard with tiebreakers based on the ratio of correct to incorrect submissions and the total submission time, favoring teams with fewer incorrect submissions.

The top 20 teams, announced on March 28th, participated in the final event on September 7th, 2023. This included an in-person CTF and a write-up session where teams presented their approach to solving a CTF challenge, with a committee evaluating presentations. Final rankings incorporated CTF scores and write-up points to determine the program's winner.

2.3 Platform

The comprehensive platform architecture is depicted in Figure 2 and incorporated the following components:

- Moodle (M. Dougiamas, 2002): Serving as an e-learning and classroom management tool, It facilitated the dissemination of training materials, including slides and recorded lessons.
- Discord (J. Citron, 2015): Employed for communication and support. After receiving invitations to join the CyberTrials Discord server, participants were randomly assigned to four-member teams upon their initial access by a Discord bot. Teams were afforded exclusive access to private Discord channels.
- CTFd: Responsible for CTF hosting and challenge provisioning.
- Instagram: Leveraged as an integral element of the quest-style CTF and for social announcements.
- WebTops: Deployed to provide all teams with a standardized, lightweight Linux environment hosted within Docker containers (S. Hykes, 2013).

Upon registration, participants received access to Moodle and Discord. Each team, through its private Discord channel, was furnished with access credentials for both the CTF platform and the dedicated Linux environment, ensuring a uniform, tailored environment conducive to their hacking activities.

3. Results

Edition

	Origin		High School year							
	NW	NE	C	S	I	1st	2nd	3rd	4th	5th
I	17%	39%	28%	6%	10%	13%	18%	29%	29%	11%
II	23%	23%	16%	28%	10%	6%	8%	43%	29%	14%

Table 1: Participants data of the two editions.

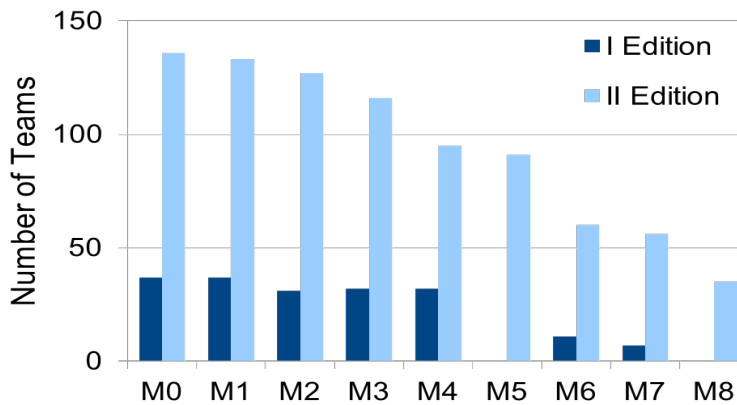
Table 1 presents aggregated data on program participants. In the second edition, participants were more evenly distributed across regions, whereas the first edition had a concentration in the northern part of the country (56%). Finally, 51% comes from a lyceum, while 49% is from a technical institute. Participation in challenges was monitored first with weekly reports on inactive members, then with warnings for inactivity, and finally excluding repeat offenders. Teams with two or fewer members were merged weekly.

Edition	Enrolled	Lecture		Challenges	
		First	Last	First	Last
I	378	189	51	181	31
II	941	557	340	440	379

Table 2: Number of participants in the two editions.

Table 2 compares attendance in last year's (I) and this year's (II) CyberTrials editions. Initial dropout matched the expected 40%, with 41% in the last edition. Lecture attendance increased from 26.9% to 61%, while challenge participation remained stable at 86%. This supports the proposition that the competitive aspect (challenges) was more engaging than lectures, as seen in the drop in participation (16.5%) after the qualification for the final.

Fig. 1
Students attendance during the two editions.



Overall, 139 teams participated, with 26 teams merging and 4 disbanding due to member defection. Figure 1 illustrates the active teams' engagement across program weeks. A notable difference between editions is the dropout rate. The first edition saw a stable participation until week 5 (M4), whereas the second edition had a more uniform decline, with the highest reduction between M5 and M6 (-23%).

Program effectiveness was assessed through pre-and post-program questionnaires. Most participants found lectures (92%) and challenges (90%) interesting and challenging (92% and 81%, respectively). Notably, 65% considered the challenges to be moderately com-

plex, aligning with participants' initial skills.

To conclude, 12,467 flags has been submitted through the CTF competition, equivalent to 9.5 flags per hour, considering day and night, weekdays and holidays.

4. Conclusion

In this article, we introduced the CyberTrials program, its gaming infrastructure, and the key outcomes observed in its initial two editions. Despite being in its second iteration, CyberTrials has demonstrated remarkable promise in terms of its appeal and efficacy in acquainting high-school girls with the field of cybersecurity. This achievement is the result of a methodical examination of the program's content and meticulous planning of its gameplay, centered around a quest Capture The Flag (CTF) format.

In conclusion, we emphasize the significance of initiatives like CyberTrials, not only for fostering interest and proficiency in cybersecurity but also for addressing the broader skill gap in computer science.

References

- D. N. Beede, T. A. Julian, D. Langdon, G. McKittrick, B. Khan, and M. E. Doms. 2011. Women in STEM: A gender gap to innovation. *Economics and Statistics Administration Issue Brief 04-11* (2011), 11.
- J. M. Bystydzienski, M. Eisenhart, and M. Bruning. 2015. High school is not too late: Developing girls' interest and engagement in engineering careers. *The Career Development Quarterly* 63, 1 (2015), 88–95.
- D. Card and A. A. Payne. 2021. High school choices and the gender gap in STEM. *Economic Inquiry* 59, 1 (2021), 9–28.
- J. Citron. 2015. Discord. Discord Inc. <https://discord.com/>
- M. Dougiamas. 2002. Welcome to the Moodle community. Moodle. <https://moodle.org>
- S. González-Pérez, R. Mateos de Cabo, and M. Sáinz. 2020. Girls in STEM: Is it a female role-model thing? *Frontiers in psychology* 11 (2020), 2204.
- S. Hykes. 2013. Docker. Docker, Inc. <https://www.docker.com/>
- J. Legewie and T. A. DiPrete. 2014. The high school environment and the gender gap in science and engineering. *Sociology of Education* 87, 4 (2014), 259–280.
- A. H. Master and A. N. Meltzoff. 2020. Cultural stereotypes and sense of belonging contribute to gender gaps in STEM. *Grantee Submission* 12, 1 (2020), 152–198.
- F. Fui-Hoon Nah, Q. Zeng, V. Rajasekhar Telaprolu, A. Padmanabhuni Ayyappa, and B. Eschenbrenner. 2014. Gamification of education: a review of literature. In *HCI in Business: First International Conference, HCIB 2014, Held as Part of HCI International 2014*, Heraklion, Crete, Greece, June 22-27, 2014. *Proceedings 1*. Springer, Berlin, Germany, 401–409.
- E. Prieto-Rodriguez, K. Sincock, R. Beretta, K. Blackmore, J. Todd, E. Wanless, S. John, and A. Giacomini. 2022. Investigating the Impact of an Outreach Intervention on Girls' STEM Identity Formation. *International Journal of Gender, Science and Technology* 14, 2 (2022), 183–206.

- C. Riegler-Crumb and K. Morton. 2017. Gendered expectations: Examining how peers shape female students' intent to pursue STEM fields. *Frontiers in psychology* 8 (2017), 329.
- H. Stoeger, X. Duan, S. Schirner, T. Greindl, and A. Ziegler. 2013. The effectiveness of a one-year online mentoring program for girls in STEM. *Computers & Education* 69 (2013), 408–418.
- M. Wang and J. L. Degol. 2017. Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational psychology review* 29 (2017), 119–140.

Autors

Gabriele Costa gabriele.costa@imtlucca.it

Gabriele Costa is Associate Professor at the SySMA Group. He received his Laurea degree in Computer Science in 2007 and his Ph.D in Computer Science in 2011. He was a member of the cybersecurity group of the Istituto di informatica e Telematica of the CNR. His appointments include a period as visiting researcher at ETH Zurich in 2016–2017. He has been co-founder of the Computer Security Laboratory (CSec) at DIBRIS. He is co-founder and CRO of Talos, a spin-off of DIBRIS focused on Cybersecurity.

His focus is on studying and applying formal methods for the automatic verification and testing of mobile and modular systems. These systems include those based on mobile software, web services, cloud and fog computing as well as the critical infrastructures and supply chains.

Serenella Valiani serenella.valiani@imtlucca.it

Serenella Valiani is a PhD student in Computer Science in the Systems Security Modelling and Analysis unit at IMT School for Advanced Studies of Lucca. She studied Mathematics at the University of Siena both at the Bachelor's and at the Master's. Currently she's studying multi-agent systems, focusing on how they can be model through formal languages based on attribute-based communication.

Progetto IS-LeGI per la didattica digitale e la formazione

Fabrizio Turchi, Francesco Romano

CNR

L'esperienza didattica fisica si sta sempre più integrando con quella virtuale. Tale processo è stato accelerato dalla pandemia che ha molto potenziato l'uso di piattaforme digitali per l'apprendimento a distanza.

Il Ministero dell'istruzione, nell'ambito della linea di investimento "Scuola 4.0" del Piano nazionale di ripresa e resilienza (PNRR) sta promuovendo la creazione di "ambienti innovativi di apprendimento" (Ministero dell'Istruzione 2022: 4).

La piattaforma IS-LeGI (Indice Semantico del Lessico Giuridico Italiano) dell'Istituto di Informatica Giuridica e Sistemi Giudiziari del Consiglio Nazionale delle Ricerche, nata per la ricerca giuridica e lessicografica sta oggi proponendosi anche quale strumento per la formazione e la didattica a vari livelli.

Lo strumento infatti, dopo essere stato usato nell'ambito del Master "La lingua del diritto" all'interno di laboratori didattici per lo studio del diritto e della lingua giuridica, viene oggi proposto come strumento per stage universitari a distanza, oltre che per laboratori interattivi progettati per studenti delle scuole medie superiori che vogliono affrontare percorsi didattici non solo in ambito giuridico, ma, più in generale, che siano interessati alla storia della lingua e del costume della nostra società.

La piattaforma IS-LeGI è stata anche proposta per una collaborazione a distanza con le studentesse e gli studenti dell'Università Suor Orsola Benincasa di Napoli con appositi stage e da due anni è usata nell'ambito dei Percorsi per le Competenze Trasversali e l'Orientamento (PCTO) che vedono coinvolte le ragazze e i ragazzi dei licei e degli istituti superiori dell'area fiorentina nello studio delle parole del diritto e dei loro significati attraverso la consultazione di documenti giuridici antichi e contemporanei. Tale acquisizione di competenze, attraverso una "metodologia didattica esperienziale, incentrata sullo studente" (Cortoni, Dattoli 2023: 52), è stata considerata particolarmente rilevante, al punto da essere menzionata nel piano didattico del Liceo statale Michelangiolo di Firenze per l'attuazione di percorsi formativi innovativi all'interno dei laboratori del Liceo.

In concreto tali percorsi didattico-formativi si servono della piattaforma IS-LeGI per implementare una banca dati attingendo alle risorse documentarie presenti nella banca dati Lessico Giuridico Italiano (LGI) che contiene oltre 900.000 schede-fonte (immagine digitale dei contesti), ottenute dallo spoglio selettivo di circa duemila testi e documenti di legislazione, dottrina, prassi e altri testi d'interesse giuridico, in grado di rappresentare la varietà delle fonti, dei rami del diritto, dei secoli (dal X al XX) e delle regioni in cui la lingua

giuridica italiana è stata ed è usata. Tale banca dati - creata con lo scopo di fornire agli utenti, non solo studiosi della lingua e del diritto, “una guida per individuare le accezioni principali riconducibili ai termini giuridici consultabili” (Mariani 2008: 241) – si pone oggi, anche come strumento didattico, per “chi vuole comprendere come cambia la nostra società, usando come parametro di riferimento l'evolversi di un linguaggio specialistico che si affida anche alla lingua comune e che dunque testimonia nel suo lessico e nella sua sintassi le trasformazioni della nostra comunità” (Romano, Cammelli 2019:104). Come noto, il linguaggio giuridico, infatti, essendo per sua natura “meno omogeneo rispetto ad altri linguaggi specialistici e tecnici” adopera termini “cui vengono attribuiti significati (accezioni) particolari” (Corasaniti 2022: 105).

Tale strumento di ricerca e di formazione unisce alle possibilità insite negli strumenti tecnologici di conservare e reperire velocemente le informazioni richieste, quella di visionare anche l'immagine del contesto da cui una data parola è tratta, andando dunque a colmare quelle difficoltà che un accesso meramente “mediato dalla tecnologia” può comportare negli studi umanistici (Tomasin 2017: 135), ma anche rendendo possibile la consultazione diretta delle fonti, in un paese, nel quale è molto difficile e faticoso giungere a contatto con i testi e i documenti (Prosperi 2021: 7).

Infatti i redattori, attraverso uno specifico software, possono associare a ciascun termine dell'Indice il significato che la parola assume nel contesto dato. In tal modo a una certa parola vengono attribuiti rilevanti metadati quali i significati e la fraseologia presente nel contesto. Le altre informazioni relative al singolo lemma (frequenze totali con indicazione delle date estreme, frequenze parziali per tipologia di fonte con date estreme, schede-fonte associate al lemma, indicazioni bibliografiche associate alle schede-fonte, sono invece a disposizione del redattore (e quindi poi anche dell'utente in fase di consultazione) nell'ambiente di redazione fornito dal software di edizione.

Per consentire un uso più sicuro della piattaforma, da parte dei redattori, è stato introdotto un sistema di autenticazione basato su protocollo sicuro TLS (Transport Layer Security) che garantisce la sicurezza di una connessione Internet e protegge i dati scambiati impedendo di leggere o modificare le informazioni trasferite. L'autenticazione permette anche di tenere traccia delle attività di ogni redattore con il fine primario di attribuire a ciascuno di loro l'elaborazione di ogni lemma, in termini di assegnazione di fraseologie, accezioni e relative immagini/schede. La redazione opera su un archivio di backend, separato da quello di frontend, visibile pubblicamente e interrogabile da chiunque. Questo permette di lavorare sull'insieme dei lemmi in maniera più sicura, senza interferire, in maniera sincrona, con i dati visibili dagli utenti. I due archivi possono essere allineati in qualunque momento utilizzando un semplice pannello di controllo (dashboard) a disposizione degli amministratori dell'archivio, dal quale si possono indicare i lemmi da visualizzare o da rendere invisibili nell'archivio frontend e anche allineare i contenuti dei due archivi. L'allineamento viene completato in maniera trasparente e autorizzato solo all'interno della rete privata dell'Istituto. Lo strumento telematico descritto ha la possibilità di supportare sia le infrastrutture di ricerca che quelle di istruzione e va ad inserirsi, a pieno titolo, in una delle quattro fasi dell'Open Science e in particolare in quella dell'Open Education, che è tesa alla realizzazione di didattica e pratiche educative aperte (Paseri 2022: 146) e deve

essere declinata come “capace di accogliere tutte le diverse concezioni di apertura, sia filosofiche sia pragmatiche, che condividono l’obiettivo di ampliare l’accesso all’educazione” (Nascimbeni 2020:18) e anche tesa a “rendere i materiali didattici ampiamente accessibili” (Nascimbeni 2020:21).

Riferimenti bibliografici

G. Corasaniti (2022), *Data science e diritto. Certezze digitali e benefici del dubbio*, Giapichelli, Torino.

C. U. Cortoni, D. Dattoli (2023), *Sapere è potere. Da Aristotele a ChatGPT, perché il futuro dipende dalla nostra formazione*, Rizzoli, Milano.

P. Mariani, (2008), «IS-LeGI: un dizionario in rete per un migliore accesso al patrimonio giuridico italiano», in *Informatica e diritto*, vol. XVII, fasc. 1-2, p. 235-244.

Ministero dell’Istruzione (2022), *Piano Scuola 4.0*, Ministero dell’Istruzione, Roma, online.

F. Nascimbeni (2020), *Open Education: OER, MOOC e pratiche didattiche aperte verso l’inclusione digitale educativa*, FrancoAngeli, Milano.

L. Paseri (2022), *Accessibilità al sapere*, in Massimo Durante, Ugo Pagallo (a cura di), *La politica dei dati. Il governo delle nuove tecnologie tra diritto, economia e società*, Mimesis, Milano-Udine.

A. Proserpi (2021), *Un tempo senza storia. La distruzione del passato*, Einaudi, Torino.

F. Romano, A. Cammelli (2019), *Banche dati di documenti giuridici per la formazione e le digital humanities*, in *selected papers della Conferenza GARR “Connecting the future”*, Politecnico di Torino, 4-5-6 giugno 2019, pp. 102-105 su <https://www.garr.it/en/garr-en/documents/selected-papers/selected-papers-conferenza-2019/4966-selected-papers-conferenza-2019-21-romano>

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

Autori



Fabrizio Turchi

Ha conseguito la laurea in Matematica presso l’Università di Firenze e attualmente è dirigente tecnologo presso l’Istituto di Informatica Giuridica e Sistemi Giudiziari (IGSG) del CNR di Firenze, dove è responsabile dei sistemi informativi. I suoi recenti interessi riguardano la rappresentazione standard dei dati e dei metadati delle prove digitali, attraverso ontologie di dominio. È stato coinvolto in molti progetti europei che hanno avuto come finalità lo scambio elettronico di prove digitali tra autorità competenti degli Stati membri dell’UE.

Francesco Romano francesco.romano@igsg.cnr.it

Laureato in Giurisprudenza è Primo ricercatore presso l’Istituto di informatica giuridica e sistemi giudiziari del CNR. Nel 1997 inizia a lavorare per il CNR dove oggi si occupa di legimatica, linguaggio giuridico, semplificazione delle comunicazioni istituzionali, legal design. Fa parte del comitato di redazione della Rivista italiana di informatica e diritto ed è uno dei responsabili scientifici del Laboratorio di ricerca su Linguaggi e Democrazia (LinDeLab) (<https://www.igsg.cnr.it/persona/francesco-romano/>).



Gemelli digitali e gestione dei dati del patrimonio culturale

Achille Felicetti¹, Franco Niccolucci²

¹PIN, Università degli Studi di Firenze, Prato, ²ARIADNE Research Infrastructure, Brussels

Abstract. Lo studio del patrimonio culturale genera enormi quantità di dati che occorre organizzare e gestire in modo efficiente per renderli utili alla ricerca. I gemelli digitali offrono un paradigma razionale per la sistematizzazione e la fruizione di questo complesso di informazioni. Il presente articolo rielabora questa tecnologia in chiave culturale e definisce un nuovo paradigma, il "gemello digitale culturale", basato su ontologie e standard internazionali in grado di descrivere organicamente reti complesse di conoscenza. Il modello risultante si propone come strumento innovativo ed efficiente per la creazione di copie digitali a tutto tondo di entità materiali e immateriali, che ne descriva la storia, il significato sociale e la valenza culturale

Keywords. Gemelli digitali, Ontologie, Standard, Beni culturali

Introduzione

Lo studio del patrimonio culturale si avvale ormai da tempo in modo sistematico di tecnologie informatiche e strumenti sofisticati per la digitalizzazione della propria documentazione. Contestualmente, assistiamo a una costante proliferazione delle informazioni digitali culturali, alla quale si accompagna però una crescente difficoltà nella loro gestione. La mole di dati generati per mezzo delle nuove tecnologie, le differenti tipologie e i diversi formati adottati, dettano infatti le condizioni per una gestione più efficiente di questa complessità, che molto spesso si traducono nella necessità di implementare paradigmi più efficienti e sviluppare infrastrutture adeguate.

In diretta prosecuzione delle tematiche discusse e delle soluzioni proposte in due precedenti lavori (Niccolucci et al. 2022; Niccolucci et al. 2023), questo articolo offre un approccio innovativo a tali problematiche, attraverso una tecnologia, quella dei gemelli digitali, ormai utilizzatissima in molti settori, ma che solo adesso muove i primi passi nel fertile terreno della gestione dei dati culturali (De Luca et al. 2022). Nelle prossime sezioni proveremo a illustrare i motivi per cui la natura peculiare di questi dati richieda soluzioni articolate, e a proporre una possibile riconfigurazione del gemello digitale come strumento privilegiato per l'organizzazione di questo tipo di documentazione.

1. Gemelli digitali e loro rilevanza

L'idea di un gemello digitale (Digital Twin) in grado di offrire rappresentazioni virtuali estremamente dettagliate e interattive di oggetti fisici, di processi e di sistemi complessi del

mondo reale non è nuova. Essa, infatti, nasce in ambito industriale, sembra per esigenze legate all'efficientamento della produzione attraverso la simulazione di scenari produttivi, l'ottimizzazione delle prestazioni, l'innovazione dei prodotti e la manutenzione predittiva. Con la sua evoluzione e grazie alle grandi potenzialità che questa tecnologia offre, il suo impiego trova posto in ambiti sempre nuovi. L'ingegneria ambientale e l'urbanistica, ad esempio, utilizzano i gemelli digitali per modellare e analizzare l'evoluzione delle città e monitorare l'impatto di decisioni urbanistiche sull'ambiente. Il gemello digitale di Singapore, esempio eminente di questo impiego, integra dati provenienti da diverse fonti (sensori, satelliti, droni, ecc.) per creare una replica virtuale della città volta a regolarne e pianificarne la crescita urbana (Caprari et al. 2022).

Nel settore dell'architettura e delle costruzioni, i gemelli digitali stanno rivoluzionando il processo di progettazione e realizzazione fornendo modelli in grado di simulare in modo accurato l'interazione di edifici e infrastrutture con l'ambiente circostante prima di essere effettivamente costruiti, rilevare precocemente problemi di progettazione e di costruzione e ridurre gli interventi di manutenzione.

Anche la ricerca scientifica sta scoprendo l'importanza di utilizzare gemelli digitali per la messa a punto di sistemi sempre più efficienti per simulare fenomeni naturali complessi e realizzare "laboratori virtuali" in grado di fornire risultati attendibili in tempi ridotti, eliminando o riducendo in tal modo i rischi e i costi della sperimentazione reale. La NASA inoltre si avvale di queste tecnologie a supporto alle missioni spaziali, come nel caso dei gemelli digitali del rover "Perseverance" e dell'elicottero "Ingenuity", realizzati per la missione Mars 2020 allo scopo simulare le loro controparti fisiche in tempo reale, garantirne il controllo da remoto e monitorarne l'interazione con l'ambiente marziano.

Di grande impatto è inoltre la proposta della Commissione Europea di creare un gemello digitale della Terra, in grado di simulare il comportamento del clima, le condizioni ambientali e i processi maggiormente impattanti sull'ambiente (quali ad esempio il riscaldamento globale) per monitorare e predire scenari futuri (European Commission 2022).

A fronte di così tanti e significativi esempi, crediamo che i tempi siano maturi per l'impiego di questa tecnologia anche sul fronte del patrimonio culturale.

2. Verso una definizione di gemello digitale culturale

Da quanto detto sopra risulta evidente che la definizione di gemello digitale varia in base alle finalità e al contesto di utilizzo per i quali viene di volta in volta impiegato. Una sua possibile applicazione all'ambito dei beni culturali richiede dunque un'attenta riflessione sulla corretta metodologia da impiegare per ottenere un livello accettabile di completezza e fedeltà all'entità reale rappresentata. Possiamo fin da subito notare che i gemelli digitali fin qui esaminati vengono tipicamente generati o prima degli oggetti reali ai quali saranno poi associati, come avviene ad esempio nella produzione predittiva di ambito industriale e nella pianificazione urbanistica, oppure contestualmente ad essi, come nel caso dei veicoli spaziali della NASA. Occorre inoltre sottolineare che nell'immaginario collettivo un gemello digitale è solitamente pensato come un modello tridimensionale digitale che sostanzialmente riproduce le caratteristiche fisiche e spaziali di un determinato oggetto reale.

Considerando il patrimonio culturale, noteremo che caratteristica comune delle entità culturali e della documentazione che le riguarda, è quella di esistere (anche da molto) prima e indipendentemente da ogni loro riproduzione o manifestazione nel mondo digitale. Il patrimonio culturale è infatti costituito da entità complesse la cui vita tipicamente si dipana lungo molteplici archi spaziotemporali. È necessario, quindi, definire la natura e l'identità degli oggetti culturali prima di formulare un opportuno paradigma per la loro rappresentazione digitale. In questa prospettiva, il “gemello digitale culturale” (Digital Heritage Twin), che è il modo in cui abbiamo denominato il paradigma che stiamo definendo, va certamente inteso come “duplicato” di un bene culturale reale, ma sarebbe riduttivo limitare tale riproduzione ai soli aspetti esteriori e materiali, tralasciando la complessa rete di informazioni di vario genere e provenienza che costituisce la parte essenziale e imprescindibile del suo “essere” oggetto culturale da un punto di vista ontologico. Le entità culturali possono inoltre anche essere immateriali, cioè prive per loro stessa natura di una qualunque forma. È il caso di eventi e manifestazioni culturali quali il Palio di Siena o il Calcio in costume fiorentino, entità che rientrano nella categoria che comunemente si definisce Intangible Heritage e che dopo la loro manifestazione nel tempo, continuano ad esistere solo sotto forma di documentazione, come testimonianza di eventi passati o formulazione di attività future.

Il complesso di questa conoscenza proviene tipicamente da documenti di vario genere, quali fonti letterarie e altra documentazione testuale, foto d'epoca, mappe, disegni, audiovisivi e altro materiale d'archivio. Di estremo interesse è anche la documentazione derivante dall'attività scientifica di analisi, osservazione, misurazione e altre investigazioni strumentali condotte su oggetti culturali. Appare evidente che è proprio il complesso di questa documentazione, opportunamente digitalizzata, a costituire il vero cuore del gemello digitale culturale e, in ultima analisi, a costituire il gemello digitale culturale stesso. Ma quali tecnologie usare per organizzare tutta questa conoscenza?

Il 3D può certamente dimostrarsi uno strumento potente e versatile per la resa e la visualizzazione degli aspetti materiali e tangibili dei beni culturali, specialmente in termini di dimensioni, forme, colori, texture, luci, ombre, ecc. In determinate condizioni è anche possibile usare il 3D come base per l'aggregazione di conoscenza di oggetti culturali complessi, come mostrato dal sistema integrato basato su 3D realizzato dal CNR di Pisa per il restauro della Fontana del Nettuno di Bologna (Apollonio et al. 2018). Tuttavia, per quanto sopra osservato, il 3D si dimostra poco adeguato alla rappresentazione di informazioni riguardo a vicende storiche, valenze artistiche, rilevanza sociale e altri elementi culturali immateriali non immediatamente traducibili in termini geometrici o visivi. Questi elementi, infatti, richiedono modalità di espressione più vicine all'ambito delle ontologie e della modellazione semantica della conoscenza.

3. L'ontologia HDT e la costruzione di un gemello digitale culturale

L'ontologia HDT (Heritage Digital Twin) è lo strumento che abbiamo sviluppato per gestire e rendere interoperabile la grande varietà di dati che, come abbiamo visto, costituisce l'ossatura del gemello digitale culturale. Il suo sviluppo nasce dall'esperienza acquisita in

importanti progetti di aggregazione e integrazione dati, quali ARIADNE (armonizzazione di dati archeologici) (Meghini et al. 2017) e 4CH (studio e preservazione di beni culturali) (4CH 2023). L'ontologia HDT è costruita come estensione del CIDOC CRM, modello di riferimento per i dati del patrimonio culturale (CIDOC CRM 2023), e propone una serie di entità intese ad esprimere in formato standard tutta la documentazione culturale e scientifica disponibile. Classi quali HC4 Intangible Aspect, HC5 Digital Representation, HC9 Heritage Activity vengono usate per modellare in modo semanticamente ricco le entità culturali e la loro documentazione, mentre specifiche proprietà, appositamente progettate, sono impiegate per esprimere le complesse relazioni esistenti con gli eventi, le persone, i luoghi e le azioni che ne compongono la storia e ne delineano il valore culturale. Grazie alla codifica semantica, tutta la conoscenza viene sistematizzata in strutture digitali complesse, dette grafi della conoscenza, solitamente ospitate su piattaforme avanzate (ad esempio, attive sul Cloud) che ne rendono possibile la consultazione e l'utilizzo tramite interfacce evolute. I grafi della conoscenza e le infrastrutture che li rendono operativi costituiscono il primo stadio del gemello digitale culturale. Eventuali operazioni di arricchimento, ad esempio con dati provenienti da sensori o altri dispositivi di monitoraggio, ne consentiranno l'estensione, mentre la messa in campo di servizi, basati su algoritmi di machine learning e di intelligenza artificiale capaci di "ragionare" sui dati archiviati, potranno effettuare simulazioni in tempo reale e generare modelli predittivi riguardo le condizioni presenti e future delle entità culturali descritte.

4. Conclusioni

In definitiva, potremmo definire il gemello digitale culturale che abbiamo fin qui delineato come la rappresentazione in formato digitale della conoscenza su entità culturali del mondo reale. L'ontologia HDT che abbiamo posto alla base del nostro lavoro ci consente infatti già da ora di implementarne quel "cuore informativo" generato dalla codifica strutturata dei dati che li rende preziosi strumenti di ricerca per studiosi ed esperti del settore. Sviluppi futuri del nostro lavoro verteranno sull'identificazione di infrastrutture ottimali per l'implementazione efficiente dei grafi della conoscenza nel Cloud, con l'obiettivo di creare spazi virtuali popolati da una molteplicità eterogenea, ma estremamente coerente e strettamente interdipendente, di gemelli digitali culturali. Questo consentirà di superare la frammentazione di sistemi dedicati a singoli oggetti, di garantire un approccio sistemico al patrimonio culturale e di fornire un contributo rilevante alla costruzione di quello "spazio comune europeo di dati del patrimonio culturale" recentemente caldeggiato dalla Commissione Europea (European Commission 2021).

Riferimenti bibliografici

- 4CH (2023) - Competence Centre for the Conservation of Cultural Heritage. <https://www.4ch-project.eu>.
- Apollonio F.I., Basilissi V., Callieri M., et al. (2018), A 3D-centered information system for the documentation of a complex restoration intervention, *Journal of Cultural Heritage*, 29, 89-99, <https://doi.org/10.1016/j.culher.2017.07.010>

Caprari G., Castelli G., Montuori M., Camardelli M., Malvezzi, R. (2022), Digital Twin for Urban Planning in the Green Deal Era: A State of the Art and Future Perspectives, Sustainability 2022, 14, 6263. <https://doi.org/10.3390/su14106263>

CIDOC CRM (2023). International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM). <http://www.cidoc-crm.org>

De Luca L., Abergel V., Guillem A., et al (2022), L'écosystème numérique n-dame pour l'analyse et la mémorisation multi-dimensionnelle du chantier scientifique Notre-Dame-de-Paris, SCAN'22 - 10e Séminaire de Conception Architecturale Numérique, Oct 2022, Lyon, France. <https://hal.science/hal-03826931>

European Commission (2022): Destination Earth.

<https://digital-strategy.ec.europa.eu/en/policies/destination-earth>

European Commission (2021): Recommendation of 10.11.2021 on a common European data space for cultural heritage.

<https://ec.europa.eu/newsroom/dae/redirection/document/80911>

Meghini C., Scopigno R., Richards J.D., et al. (2017), ARIADNE: A Research Infrastructure for Archaeology. Journal on Computing and Cultural Heritage, 10(3), pp.1–27

Niccolucci F., Felicetti A., Hermon S. (2022), Populating the Data Space for Cultural Heritage with Heritage Digital Twins, Data, 7, 105. <https://doi.org/10.3390/data7080105>

Niccolucci F. et al. (2023), The Heritage Digital Twin: a bicycle made for two. The integration of digital methodologies into cultural heritage research, Open Research Europe, 3,64. <https://doi.org/10.12688/openreseurope.15496.1>

Autori

Achille Felicetti achille.felicetti@pin.unifi.it

Achille Felicetti è un archeologo e ricercatore presso il laboratorio VAST-LAB del PIN di Prato. Si occupa della definizione e dell'applicazione di standard e modelli concettuali in ambito umanistico. Coordina diversi gruppi di sviluppo di strumenti per la gestione della documentazione dei beni culturali. Collabora alla definizione di ontologie quali CRMarchaeo e CRMtex, estensioni del modello ontologico CIDOC CRM. Coordina lo sviluppo di basi semantiche della conoscenza in varie iniziative internazionali, fra le quali ARIADNE Research Infrastructure e 4CH.

Franco Niccolucci franco.niccolucci@gmail.com

Franco Niccolucci è l'attuale presidente di ARIADNE Research Infrastructure AISBL, l'organizzazione che nasce dall'omonimo progetto – di cui è stato coordinatore – e raccoglie oltre 30 istituzioni di ricerca europee operanti nel campo delle tecniche digitali in archeologia e i beni culturali. Conserva inoltre la sua posizione di direttore del laboratorio VAST-LAB presso il PIN.

Play INAF: risorse digitali per la didattica dell'astronomia

Maura Sandri¹, Rino Bandiera², Silvia Casu³, Laura Daricello⁴, Federico Di Giacomo¹, Mariachiara Falco⁵, Maria Teresa Fulco⁶, Silvia Galletti¹, Giuliana Giobbi⁷, Laura Leonardi⁴, Marco Malaspina¹, Claudia Mignone⁷, Barbara Olmi², Serena Pastore⁸, Agatino Rifatto⁶, Rachele Toniolo⁹, Ginevra Trinchieri¹⁰

¹INAF – OAS Bologna, ²INAF – O.A. Arcetri, ³INAF – O.A. Cagliari, ⁴INAF – O.A. Palermo, ⁵INAF – O.A. Catania, ⁶INAF – O.A. Capodimonte, ⁷INAF – Sede Centrale, Roma, ⁸INAF – O.A. Padova, ⁹INAF – IRA Bologna, ¹⁰INAF – O.A. Brera

Abstract. Play INAF è la piattaforma di didattica innovativa dell'Istituto nazionale di astrofisica, che quest'anno ha ricevuto il premio CLASSified Award 2023 come uno dei 100 migliori prodotti del 2022 nel settore dell'innovazione tecnologica e digitale. Le risorse presenti sulla piattaforma riguardano attività di coding, tinkering, making, robotica educativa, realtà aumentata e virtuale. Il focus del contributo è sulle risorse digitali disponibili sulla piattaforma, pensate per permettere agli studenti di acquisire abilità di programmazione e di problem solving, e nel contempo imparare concetti di astronomia, divertendosi

Keywords. Coding, pensiero computazionale, problem solving, competenze digitali, STEM

Introduzione

Play INAF è la piattaforma di didattica innovativa dell'Istituto Nazionale di Astrofisica (INAF) che quest'anno ha ricevuto il premio CLASSified Award 2023 come uno dei 100 migliori prodotti del 2022 nel settore dell'innovazione tecnologica e digitale, nella categoria “Educare i figli”. Il filo conduttore di tutte le risorse proposte sulla piattaforma è la didattica dell'astronomia e della fisica. Tutte le attività – rivolte a famiglie, studenti e insegnanti, e suddivise in otto categorie: coding, robotica educativa, making, tinkering, laboratori hands-on, realtà virtuale, realtà aumentata e giochi – sono originali, gratuite, liberamente accessibili, disponibili in più lingue (italiano, inglese, francese e tedesco). Il presente contributo riguarda in particolare le risorse digitali presenti sulla piattaforma, attraverso le quali gli studenti possono sviluppare il pensiero computazionale, le abilità di problem solving e le competenze digitali. Tra queste, entreremo nel dettaglio di tre attività che hanno riscosso notevole successo durante i maggiori festival della scienza in Italia: il Cody Maze astrofisico, la Code Hunting Game astrofisica internazionale e la escape room “A cavallo di un fotone”.

1. Cody Maze astrofisico

Il Cody Maze astrofisico è un labirinto virtuale nel mondo reale che propone sfide di coding e quiz di astronomia, astrofisica ed esplorazione spaziale (Sandri et al. 2023). Il gioco consiste in una scacchiera 5 x 5 di codici QR che, attraverso l'interazione con un bot Telegram, indirizzano i giocatori lungo labirinti sempre diversi. Durante le varie tappe del gioco vengono proposte sequenze di istruzioni di complessità crescente che introducono i concetti base della programmazione. Per ricevere la sequenza di istruzioni che porta al codice QR successivo occorre rispondere correttamente a un quiz il cui tema dipende dalla casella del labirinto: dai pianeti e corpi minori del Sistema solare a stelle e galassie; dalla storia dell'esplorazione spaziale alle particelle che popolavano l'universo primordiale. Gli enigmi di coding e i quiz sono rivolti a studenti a partire dalla scuola secondaria di primo grado, e a un pubblico adulto. Il gioco, realizzato da INAF in collaborazione con l'Università di Urbino e Digit Srl, è stato lanciato in anteprima al Festival della Scienza di Genova del 2021 e riproposto al Festival delle Scienze di Roma dello stesso anno. Nel 2022, per la Notte Europea dei Ricercatori, è stato proposto al pubblico da ricercatori e tecnologi di molte sedi INAF (Bologna, Cagliari, Roma, Trieste, Napoli, Palermo), coinvolgendo oltre 2000 partecipanti.

Fig. 1
Installazioni
del Cody Maze
astrofisico a
Padova, Geno-
va, Matera e
Cagliari.



2. Code Hunting Game astrofisica internazionale

La Code Hunting Game astrofisica internazionale è un'impresa unica nel suo genere, ideata e realizzata nel 2021: cinque lingue, sei continenti, oltre trecento luoghi da scoprire (Mignone et al. 2022). Si tratta di una caccia al tesoro virtuale che accompagna chi gioca alla scoperta di osservatori astronomici attualmente operativi, dai più grandi e famosi ad altri più piccoli o meno noti, accanto a luoghi iconici e personaggi illustri nella storia dell'astrofisica, ma anche rivelatori di particelle e onde gravitazionali, basi di lancio e siti di ritrovamento di meteoriti caduti sulla Terra. Interagendo con un bot Telegram, i parteci-

panti devono risolvere una serie di quesiti di coding e, tra una sfida e l'altra, risolvere quiz legati all'astronomia, muovendosi sulla mappa del mondo e saltando da un continente all'altro alla ricerca dei luoghi da cui si osserva, si studia, si cerca di comprendere e si racconta la struttura, la storia, l'origine e il funzionamento dell'universo. Può essere giocata sia a casa che in classe, da soli oppure in squadra, dividendosi i compiti e collaborando nella risoluzione delle varie sfide incontrate. Si tratta di uno strumento educativo potente che apre la mente, sviluppa l'intuizione, stimola con curiosità a conoscere nuovi luoghi, avvalendosi di messaggi, prove e vere e proprie sfide di programmazione (Klopfenstein et al. 2018).

3. Escape room "A cavallo di un fotone"

"A cavallo di un fotone" è la escape room di INAF nella quale i partecipanti si immedesimano in una particella di luce (fotone) per emergere dalle profondità del Sole e dirigersi verso i pianeti del Sistema solare, fino a raggiungere i telescopi sulla Terra. Per superare le diverse prove – che includono una serie di attività sviluppate utilizzando l'approccio del coding unplugged (ovvero la programmazione senza l'uso di computer) – occorre mettere in campo abilità logiche e di problem solving, programmando piccoli robot e risolvendo enigmi di fisica e astronomia, per trovare le "chiavi" che permettono di aprire congegni criptati (cryptex) e lucchetti, all'interno dei quali si trovano altre sfide.

Questa escape room è stata presentata in anteprima al Festival della Scienza di Genova del 2022, dove è stata accolta molto positivamente, coinvolgendo oltre 1900 partecipanti tra i quali più di 40 classi di tutti i gradi scolastici, dalle scuole primarie fino alle secondarie di secondo grado. L'attività prevede sfide di natura molto diversa (programmare robot, interpretare codici, colorare, cercare indizi, rispondere a quiz, risolvere problemi combinando diversi oggetti in modo creativo) in modo da incoraggiare di volta in volta la partecipazione di studenti con diverse abilità, coinvolgendo anche chi non ha una propensione spiccata verso la fisica e l'astronomia. Il design del gioco indirizza i diversi partecipanti (o squadre) su percorsi paralleli, raggiungendo diversi pianeti, e dunque a competere per completare tutte le sfide e a "uscire" nel minor tempo possibile. Nello spirito dell'Open Education, questa esperienza può essere replicata in altri contesti quali scuole, musei, centri visite, planetari, ludoteche.

Fig. 2
Escape room "A cavallo di un fotone"



4. Altre risorse di coding

Rimanendo nell'ambito dello sviluppo del pensiero computazionale tramite il coding, la piattaforma propone anche attività realizzate servendosi di Arduino, una piattaforma hardware composta da una serie di schede elettroniche dotate di un microcontrollore, con la quale si possono realizzare semplici dispositivi che utilizzano sensori e attuatori, alla quale è abbinato un ambiente di sviluppo per la programmazione del microcontrollore. Tutto il software è libero, così come lo sono tutte le attività riportate su Play INAF che permettono, ad esempio, di realizzare una bilancia gravitazionale con Arduino, di studiare la forza elettromotrice indotta, di accendere le costellazioni e di capire come vengono scoperti nuovi pianeti con il metodo dei transiti. Con solo un tablet o un pc è possibile imparare a programmare utilizzando Scratch, un ambiente di programmazione visuale sviluppato dal gruppo di ricerca Lifelong Kindergarten presso il Multimedia Lab del MIT di Boston, particolarmente adatto per insegnare le basi della programmazione agli studenti, anche giovanissimi. Nella categoria coding con Scratch della piattaforma sono presenti attività che spiegano come fare un astro-documentario, permettono di scoprire quanto pesiamo sui vari pianeti nel Sistema solare, perché i sassi affondano e le navi galleggiano, comprendere la fisica alla base delle Leggi di Keplero e avvicinarsi ai modelli scientifici degli antichi, come ad esempio il moto planetario secondo il modello Tolemaico.

5. Conclusioni

Da quando è stata pubblicata, nel settembre del 2020, Play INAF conta oltre 55mila utenti. Negli ultimi 12 mesi i nuovi utenti sono stati 29mila, distribuiti in tutto il mondo, a testimonianza del fatto che la piattaforma è, di fatto, una realtà internazionale, al servizio della comunità. Ricercatori e tecnologi di INAF pubblicano attività didattiche con continuità, poiché sono convinti che in un'epoca in cui i settori della scienza, della tecnologia, dell'ingegneria e della matematica sono in continua evoluzione e costituiscono la base di una serie di lavori fondamentali della società, sia fondamentale contribuire a far sì che i giovani acquisiscano e sviluppino queste competenze e abilità.

Riferimenti bibliografici

M. Sandri, C. Mignon, A. Bogliolo, R. Bandiera, D. Coero Borga, S. Casu, M. T. Fulco, G. Giobbi, L. C. Klopfenstein, S. Galletti, L. Leonardi, M. Malaspina, S. Ricciardi, R. Toniolo, G. Trinchieri (2023), The International Astrophysical Online Code Hunting Game and the Astrophysical Cody Maze, Mem. S.A.It. Vol. 75, 282

Mignone, C., Sandri, M., Bandiera, R., Bogliolo, A., Casu, S., Fulco, M. T., Giobbi, G., Klopfenstein, L. C., Leonardi, L., Malaspina, M., Toniolo, R. & Trinchieri G. (2022), The first international online Astrophysical Code Hunting Game. GARR Conference Proceedings 13

Klopfenstein, L. C., Delpriori, S., Paolini, B. D., & Bogliolo, A. (2018), Code Hunting Games: A Mixed Reality Multiplayer Treasure Hunt Through a Conversational Interface.

Internet Science. INSCI 2017. Lecture Notes in Computer Science, vol 10750. Springer, Cham.

Autori

Maura Sandri



Tecnologa e giornalista dell'INAF. Si occupa della progettazione e caratterizzazione di ottiche per osservazioni a microonde e nella banda radio, nonché di didattica e divulgazione. È coordinatrice del working group per lo sviluppo di risorse di coding e robotica educativa per la scuola, ideatrice e amministratrice della piattaforma play.inaf.it. Fa parte della redazione di Media INAF. È direttrice responsabile della rivista Universi, house organ di INAF.

Rino Bandiera

Astronomo dell'Istituto Nazionale di Astrofisica, Osservatorio Astrofisico di Arcetri (Firenze). Principali campi di ricerca: resti di supernove, pulsar, nebulose associate a pulsar, accelerazione di particelle, fenomeni di alta energia nel mezzo interstellare. All'interno di Play INAF si interessa di progetti legati al coding.



Silvia Casu



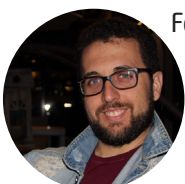
Ricercatrice presso il Gruppo di Astrochimica e Astrobiologia dell'INAF-Osservatorio Astronomico di Cagliari. Responsabile dell'Area Comunicazione e Divulgazione dell'OAC.

Laura Daricello

Tecnologo presso INAF, è Responsabile del Servizio Comunicazione e Attività per il Pubblico dell'INAF Osservatorio Astronomico di Palermo, coordinatore nazionale delle attività legate alla realtà virtuale e aumentata per la diffusione dell'astrofisica e P.I. del Progetto PRIN INAF "Virtual Reality and Augmented Reality for Science, Education and Outreach". Collabora con la redazione di Edu Inaf, il magazine di didattica e divulgazione dell'Inaf, e con Play Inaf, la piattaforma di didattica innovativa dell'Inaf.



Federico Di Giacomo



Laureato in astrofisica e cosmologia presso l'università di Bologna. Dal 2015 lavora presso l'INAF nell'ambito della didattica e divulgazione. Attualmente possiede un contratto presso l'INAF di Bologna dove si occupa di progettare attività di public engagement per il progetto CTA+ mediante l'utilizzo di tecnologie emergenti. Oltre a questo, è impegnato nel campo della valorizzazione del patrimonio scientifico, oltre che a dedicarsi a varie attività di outreach.

Mariachiara Falco

Assegnista di ricerca presso l'INAF-Osservatorio Astrofisico di Catania si occupa di ricerca nel campo della fisica solare ed è vice responsabile dell'Ufficio Rapporti con il Pubblico e Divulgazione dell'INAF-OAcT. Organizza eventi di divulgazione sul territorio con le scuole e il pubblico generico, è redattrice del magazine online di didattica e divulgazione EduINAF e collabora con il gruppo di Play INAF.





Maria Teresa Fulco

Dipendente INAF. Membro dell'International Astronomical Union. Comunicatore della scienza, si dedica in particolare ai bambini di età scolare e prescolare. Promuove la cultura astronomica nelle scuole e organizzando eventi aperti al pubblico. Elabora materiale didattico con i gruppi nazionali Edu INAF e Play INAF.

Silvia Galletti

Tecnologa dell'INAF, collabora al progetto sul monitoraggio di satelliti artificiali e space debris con osservazioni ottiche da terra nell'ambito del consorzio europeo EU-SST. Membro del gruppo Play-INAF per lo sviluppo di risorse di coding e robotica educativa per la scuola. Responsabile della sede interregionale Emilia Romagna e Toscana dei Campionati Italiani di Astronomia. Si occupa della didattica e divulgazione dell'astronomia per il pubblico alla Stazione Astronomica di Loiano.



Giuliana Giobbi

Laureata in Lingue Straniere Moderne presso l'Università La Sapienza di Roma, ha conseguito un Dottorato di Ricerca presso l'Università di Glasgow (UK) in Letteratura Inglese e Comparata. Tornata in Italia, lavora da molti anni in INAF, prima presso INAF-OAR e da due anni presso la Sede Centrale: al momento è responsabile della Segreteria di Presidenza e continua ad occuparsi di divulgazione, di traduzione e di organizzazione di eventi nazionali ed internazionali.

Laura Leonardi

Giornalista pubblicista e assegnista di ricerca, collabora con l'Ufficio comunicazione dell'INAF e lavora presso l'Osservatorio Astronomico di Palermo (INAF OAPa) nell'ambito dello sviluppo di tecnologie innovative e prodotti multimediali con applicazioni di realtà virtuale, realtà aumentata e di computer grafica per la diffusione della cultura scientifica e la valorizzazione del patrimonio storico-scientifico.



Marco Malaspina

Giornalista professionista, lavora all'Ufficio comunicazione dell'Inaf, dove dirige la testata online Media Inaf. In passato ha insegnato al Master in comunicazione della scienza della Sissa, ha condotto il programma radiofonico Pigreco Party, è stato tra i fondatori di Formicablu e ha collaborato per anni come freelance con numerose testate, soprattutto per le pagine di salute del settimanale Oggi. È l'autore di La scienza dei Simpson (Sironi, 2007).

Claudia Mignone

Astrofisica e comunicatrice scientifica, lavora come tecnologa nell'area comunicazione e divulgazione presso la sede centrale dell'INAF a Roma. In passato, ha lavorato per dieci anni alla comunicazione delle missioni scientifiche dell'Agenzia Spaziale Europea (ESA) nei Paesi Bassi.



Barbara Olmi

Ricercatrice INAF dal 2020, si occupa di astrofisica teorica delle alte energie. Diplomata

alla scuola internazionale di comics, per passione si dedica a progetti di grafica per Play inaf.

Serena Pastore

Ingegnere elettronico, lavora presso INAF Padova come tecnologo ICT, per molti anni nelle infrastrutture di calcolo a supporto della ricerca astrofisica, ma coinvolta in attività di terza missione per disseminazione e trasferimento tecnologico. Attualmente docente a contratto presso Università di Ferrara, dal 2014 lavora come esperto indipendente nella valutazione e monitoraggio in programmi comunitari di ambito ricerca e innovazione ed educativo e formativo.



Agatino Rifatto

Astronomo presso l'INAF/Osservatorio Astronomico di Capodimonte, Napoli (OACNA) dal 1990. Laureato in Astronomia presso Università degli Studi di Padova. Componente del Consiglio Direttivo e Tesoriere della Società Astronomica Italiana (SAIt). Professore a contratto di Astrofisica all'Università di Salerno dal 2002 al 2008. Correlatore di tesi di Laurea e di dottorato presso l'Università Federico II di Napoli e l'Università di Padova. Responsabile delle attività D&D dell'INAF/OACNA. Organizzatore e Direttore di Corsi di Formazione per docenti e Scuole Estive per studenti della SAIt.



Rachele Toniolo

Studentessa di dottorato in Astrofisica presso l'Università di Bologna. Si occupa dello studio e sviluppo di nuove risorse didattiche inerenti la radioastronomia, sfruttando diverse metodologie innovative, con particolare attenzione all'apprendimento tramite il gioco (game based learning).



Ginevra Trinchieri

Astrofisica interessata allo studio delle galassie e degli ammassi di galassie, delle loro proprietà e della loro evoluzione. Si occupa anche di comunicazione della scienza e della salvaguardia del patrimonio storico scientifico dell'INAF. Rappresentante Nazionale presso International Astronomical Union.



Il Sistema di Gestione della Sicurezza (SGSL) dell'INAF e integrazione con il database dell'anagrafica dei dipendenti

Vito Conforti¹, Francesca Romana Porta¹, Elisabetta Bartone¹ Giuseppina Antico¹, Angelo Boccaccini¹, Sabrina Ciprietti¹, Gianluigi Deiana¹, Nicola Di Cicco¹, Stefano Giovannini¹, Davide Inamo¹, Jacopo Lenzi¹, Giovanni Liggio¹, Sergio Mariotti¹, Francesco Massaro¹, Barbara Neri¹, Giovanni Occhipinti¹, Mario Pepe¹, Francesco Perrotta¹, Pier Raffaele Platania¹, Irene Zago²

¹Istituto Nazionale di Astrofisica (INAF), ²AGE web solutions

Abstract. L'Istituto Nazionale di Astrofisica (INAF) è il principale Ente di ricerca italiano per lo studio dell'Universo. Più di 1500 lavoratori operano in numerose strutture di ricerca distribuite sul territorio nazionale. Le strutture ospitano facility osservative, laboratori, officine con strumentazioni e sostanze che espongono il personale a determinati rischi. L'INAF ha implementato una struttura organizzativa del servizio di prevenzione e sicurezza al fine di garantire sicurezza, salute e benessere di lavoratori e lavoratrici. Per ottimizzare i processi del servizio, INAF ha acquistato, dalla stessa ditta, un software di gestione per la sicurezza sul lavoro e un software per la valutazione dei rischi, progettati per garantire completa interoperabilità dei dati tra i due sistemi, e implementano identico data model per l'anagrafica del lavoratore. In questo articolo presentiamo integrazione e interoperabilità di dati tra software INAF e software commerciali, valutazioni e sviluppi futuri

Keywords. rischi, prevenzione, benessere, software, interoperabilità

Introduzione

L'Istituto Nazionale di Astrofisica (INAF) promuove, realizza e coordina, anche nell'ambito di programmi dell'Unione Europea e di Organismi Internazionali, attività di ricerca nei campi dell'astronomia e dell'astrofisica, sia in collaborazione con le Università che con altri soggetti pubblici e privati, nazionali, internazionali ed esteri. INAF inoltre progetta e sviluppa tecnologie innovative e strumentazione d'avanguardia per lo studio e l'esplorazione del Cosmo. Il servizio di prevenzione e sicurezza dell'INAF fornisce assistenza e supporto al fine di definire e mettere in atto tutte le procedure che consentono di operare in sicurezza nell'ambito delle attività istituzionali dell'Ente e della loro diffusione sul territorio, ed è costituito da un servizio di staff alla Direzione Generale denominato Servizio di Prevenzione e Sicurezza sul Lavoro (SPSL) e dai Responsabili del Servizio Prevenzione e Protezione (RSPP) delle strutture di ricerca.

Il SPSL e gli RSPP hanno una funzione di consulenza per il datore di lavoro e in particolare nella individuazione dei fattori di rischio, valutazione dei rischi, individuazione delle misure di sicurezza e per la salubrità degli ambienti di lavoro, elaborazione di misure preventive e protettive previste nel Documento di valutazione dei rischi (DVR), elaborazione di procedure di sicurezza per le varie attività, preparazione del piano di formazione e informazione.

Il servizio di prevenzione e sicurezza dell'INAF gestisce una ingente quantità di dati e informazioni complesse e in costante aggiornamento anche in relazione alle proprie attività istituzionali, alla articolazione delle strutture di ricerca, ai cambiamenti dei livelli di rischio delle attività e il monitoraggio dei processi e alla normativa di settore in continua evoluzione.

Per ottimizzare la gestione e l'aggiornamento dei dati e delle informazioni, anche ai fini della unicità del sistema documentale e della gestione dei dati, evitando la duplicazione di registrazioni comuni ai vari ambiti e facilitando la comunicazione interna attraverso la condivisione delle informazioni l'Ente ha deciso di adottare un Sistema di Gestione della Sicurezza (SGSL) (Tullini 2010), e un Software per la Valutazione dei Rischi (SVR) ai sensi dell'articolo 30 del Decreto Legislativo del 16 aprile 2008, numero 81, e successive modifiche e integrazioni (Amato, Di Fiore 2016), per i seguenti principali motivi:

- investimento in termini di tempo e di risorse, risparmi sulle procedure, sulla formazione, prevenzione di errori/conflicti tra normative indipendenti;
- miglioramento continuo delle condizioni di salute e la sicurezza dei lavoratori, nell'ottica dell'efficacia e dell'efficienza globale e del benessere lavorativo, adottando uno strumento che consente di tenere sotto controllo i processi e le attività di tutto l'Ente;
- individuazione di nuovi approcci strategici grazie all'analisi dei processi globali, in ottica di impatto ambientale e in termini di riduzione degli incidenti e degli infortuni, con conseguente aumento dell'efficienza delle attività dell'Ente e della sua produttività.

L'Ente promuove, tra l'altro, numerose iniziative di grande interesse e valore scientifico culturale presso le proprie strutture (sale conferenze, laboratori, telescopi, planetari...) rivolte a un pubblico molto vasto che comprende anche studenti di scuole di vario ordine e grado. Queste attività rendono necessario garantire un elevato livello di sicurezza degli edifici e degli impianti per la protezione e salvaguardia delle persone (personale, visitatori, collaboratori) e del patrimonio immobiliare.

Il sistema di gestione della sicurezza, piattaforma unica per tutto l'Ente che si basa su un algoritmo di calcolo sulle metodologie descritte all'interno della ISO 31000, riguarda anche gli aspetti delle manutenzioni del patrimonio immobiliare consentendo di tenere traccia di tutti i processi manutentivi effettuati (siano essi ordinari o straordinari), e dando la possibilità di gestire le manutenzioni complesse, i carichi di lavoro, le assegnazioni, il magazzino dei pezzi di ricambio, i costi di manutenzione, promuovendo l'efficienza e l'efficacia dei modelli di gestione della sicurezza, anche in termini di qualità.

Inoltre l'uso di una piattaforma unica per tutto l'Istituto consente di gestire e ottimizzare l'insieme dei processi di tutte le strutture di ricerca e di ridurre il numero di apparecchi-

ture delle singole strutture necessarie alla memorizzazione, gestione dinamica e backup dei dati. Tra i vantaggi, l'efficienza dei processi e la riduzione dei device favorisce la sostenibilità nel campo dell'Information and Communication Technology (ICT).

1. Interoperabilità dei dati tra SGSL, SVR e anagrafica INAF

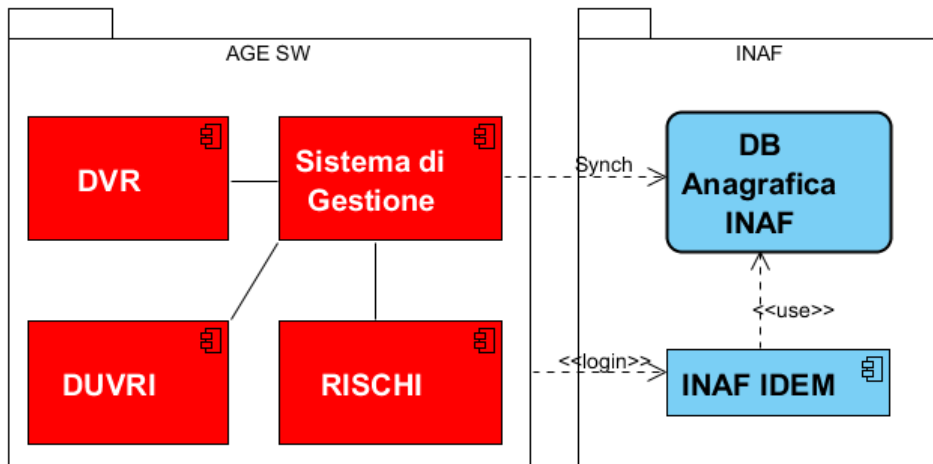
I software SGSL e SVR adottati dall' INAF sono software web based, acquisiti con licenza commerciale da Age Web Solution, che si integrano con due principali servizi:

- INAF IDEntità Management IDEM per accesso federato;
- Database Anagrafica INAF

L'identità IDEM fornisce un servizio di Single Sign-On (SSO) per tutto il personale, indipendentemente dal ruolo, e consente di accedere con le proprie credenziali a SGSL e SVR. L'integrazione di SGSL e SVR con il database Anagrafica INAF consente di avere un aggiornamento costante del personale. Questa funzionalità è utile per poter gestire attività quali: sorveglianza sanitaria, formazione, informazione e addestramento in tematiche inerenti alla sicurezza, dispositivi di protezione individuale, gestione delle manutenzioni e delle segnalazioni, fattori di rischio e gestione del rischio stesso. Il software è stato configurato opportunamente con diversi profili e aree di competenza per poter gestire in modo efficace le attività del servizio sia a livello di singola struttura sia a livello di Ente.

Come illustrato in fig. 1 il sistema di gestione si sincronizza con il database anagrafica INAF, per registrare nuovi utenti, aggiornare utenti esistenti o cancellare utenti cessati. Il sistema di gestione consente di propagare questi aggiornamenti agli altri componenti software DVR (Documento di Valutazione dei Rischi), DUVRI (Documento Unico di Valutazione dei Rischi Interferenti) e il modulo rischi. Allo stesso modo l'autenticazione per l'accesso al software di gestione fa uso del sistema di SSO di INAF.

Fig. 1
INAF e AGE Web
Solution data
flow



2. Considerazioni per incrementare l'interoperabilità

L'interoperabilità permette alle persone di collaborare in piena sinergia tra loro, di condividere i dati in maniera veloce, efficace e sicura, riducendo al minimo le possibilità di errori. I vantaggi dell'interoperabilità dei dati sono ormai noti a tutti e lo vediamo nella vita di

tutti i giorni. Un esempio molto noto a livello nazionale è il Sistema Pubblico di Identità Digitale (SPID) (Agenzia per l'Italia digitale 2017) che ci consente di accedere a numerosi servizi senza l'onere di fare mille registrazioni e inserire o aggiornare più volte gli stessi dati. Inoltre l'Agenzia per l'Italia Digitale (AGID) ha emanato le linee guida sull'infrastruttura tecnologica (Agenzia per l'Italia digitale 2021), tecnologie e standard tramite API (Agenzia per l'Italia digitale 2022), e interoperabilità tecnica delle Pubbliche Amministrazioni (PA) (Agenzia per l'Italia digitale 2023). Ci sono quindi tutti i presupposti per incrementare l'interoperabilità dei dati tra i sistemi. Sulla base della nostra esperienza occorre fare alcune considerazioni relativamente ai vincoli:

- giuridici: software acquisiti in outsourcing, come nel caso di SGL e SVR sono concessi con specifiche licenze d'uso, quindi non si ha necessariamente l'accesso alle interfacce per poter accedere facilmente al dato. Solitamente software open source, GPL consentono maggiori flessibilità e accessi ai dati a discapito di aggiornamenti di funzionalità, informazioni, sicurezza, e supporto che invece sono solitamente garantiti da software commerciali.
- Tecnici: la persistenza e gestione del dato è garantito dal DBMS (Data Base Management System). Esistono diversi tipi di DBMS (relazionali, orientati ai documenti, orientati agli oggetti, gerarchici), con API sviluppate in diversi linguaggi di programmazione. Riteniamo necessario considerare questi vincoli in fase di analisi dei requisiti, e di progettazione per l'acquisizione di nuovi software, siano essi sviluppati internamente o reperiti da terzi.

3. Conclusioni

Il sistema di gestione della sicurezza e il Software per la Valutazione dei Rischi integrati con l'IDEntità Management IDEM per accesso federato e il Database Anagrafica INAF comportano, dunque, numerosi vantaggi relativamente a molteplici aspetti, e consentono di gestire e unificare processi e dati delle strutture INAF in materia di prevenzione e sicurezza sul lavoro. I risultati attesi dell'uso di una piattaforma unica per tutto l'Ente riguardano la gestione, in modo coordinato, dell'insieme dei processi, l'ottimizzazione degli interventi, delle azioni e delle procedure di controllo in materia di prevenzione e sicurezza riducendo, contestualmente, il numero di strumenti e di apparecchiature in dotazione alle singole strutture attualmente utilizzate per la memorizzazione, la gestione dinamica, l'analisi, conservazione e il backup dei dati. L'interoperabilità dei dati tra questi servizi ha un impatto significativo sui processi gestionali relativi al servizio protezione e prevenzione, promuovendo l'efficienza e l'efficacia dei modelli di gestione della sicurezza, anche in termini di qualità. I sistemi di IDEntity Management IDEM e il Database Anagrafica INAF sono ampiamente utilizzati anche per tutti gli altri servizi INAF (gestione delle risorse umane, del cartellino, delle missioni, ecc.). In futuro riteniamo doveroso valutare l'impatto costi/benefici per incrementare l'interoperabilità tra tutti i servizi, sia quelli in essere sia i nuovi.

Riferimenti bibliografici

Agenzia per l'Italia digitale (2017), SPID Sistema Pubblico di Identità Digitale General Information, https://eidas.agid.gov.it/spid/SPID-General_information.pdf

Agenzia per l'Italia digitale (2021), Linee Guida sull'infrastruttura tecnologica della Piattaforma Digitale Nazionale Dati per l'interoperabilità dei sistemi informativi e delle basi di dati, https://www.agid.gov.it/sites/default/files/repository_files/lg_infrastruttura_interoperabilita_pdnd.pdf

Agenzia per l'Italia digitale (2022), Linee Guida Tecnologie e Standard per la sicurezza dell'interoperabilità tramite API dei sistemi informatici, https://www.agid.gov.it/sites/default/files/repository_files/linee_guida_tecnologie_e_standard_sicurezza_interoperabilit_api_sistemi_informatici.pdf

Agenzia per l'Italia digitale (2023), Linee Guida sull'interoperabilità tecnica delle Pubbliche Amministrazioni, https://www.agid.gov.it/sites/default/files/repository_files/linee_guida_interoperabilit_tecnica_pa.pdf

G. Amato, F. Di Fiore (2016), Testo Unico sulla salute e sicurezza sul lavoro.

P. Tullini (2010), I sistemi di gestione della prevenzione e della sicurezza sul lavoro, Giornale di diritto del lavoro e di relazioni industriali. Fascicolo 3, 2010, Milano, pp 1-12.

Autori



Vito Conforti vito.conforti@inaf.it

Tecnologo presso la sede INAF di Bologna dal 2012. Laureato in Informatica nel 2009 ha acquisito skills lavorando in ambito privato, e all'Università di Bologna prima di approdare in INAF e fornire il suo contributo in progetti da terra e missioni spaziali. Vito si occupa di progettazione, sviluppo, test e management di software. Poco prima della pandemia è stato nominato RSPP per la sede di Bologna ed ha contribuito a minimizzare il rischio di contagio anche attraverso competenze informatiche.

Francesca Romana Porta francesca.porta@inaf.it

Tecnologa presso l'INAF, Architetto, si occupa di lavori pubblici, sicurezza sul lavoro, patrimonio immobiliare, sostenibilità energetica e ambientale, con incarichi di coordinamento e responsabilità a livello centrale e nazionale. Partecipa a Gruppi di Lavoro, Tavoli Tecnici, Commissioni. E' componente del Comitato Unico di Garanzia (CUG) INAF. Ha redatto come unica autrice/progettista e in team progetti, rapporti tecnici, atti, linee guida, manuali, protocolli, relazioni e note tecniche.



Elisabetta Bartone elisabetta.bartone@inaf.it

In servizio presso la Direzione Generale INAF si occupa di lavori pubblici, patrimonio immobiliare, sicurezza e sostenibilità energetica e ambientale con funzioni di supporto tec-

nico, coordinamento, monitoraggio a livello nazionale. È Preposta per la sicurezza, Coordinatrice Generale del Piano di Emergenza, Project Manager RUP. Collabora con Tavoli Tecnici Permanenti, Gruppi di lavoro, commissioni, comitati dell'Ente. Ha redatto come autrice e co-autrice numerosi progetti e pubblicazioni.

Giuseppina Antico giuseppina.antico@inaf.it

In servizio presso l'Osservatorio Astrofisico di Torino. Si è occupata del personale (gestione presenze, prese di servizio, creazione database personale, rapporti con altri Enti); attualmente fa parte dell'ufficio tecnico (gestione manutenzioni ordinarie e straordinarie, capitolati tecnici, ricerche di mercato, responsabile del procedimento ordini, attivazione convenzioni Consip, protocollo documenti, collaudo fatture, creazione database inventario). E' inoltre il ruolo di RSPP della Struttura.



Angelo Boccaccini angelo.boccaccini@inaf.it

Angelo Boccaccini è un IV livello CTER presso lo IAPS. Ha iniziato la sua esperienza lavorativa dal 2000 presso il laboratorio di Planetologia, occupandosi della progettazione meccanica con integrazione di sistemi di controllo attraverso il linguaggio di programmazione sviluppato dalla National Instruments per la realizzazione di sistemi automatici di calibrazione, misura e/o controllo.

Sabrina Ciprietti sabrina.ciprietti@inaf.it

In servizio presso l'Osservatorio Astronomico d'Abruzzo dal 2013, dove si occupa in particolare dell'Ufficio del Personale, della segreteria di Direzione e della gestione documentale. Ricopre inoltre il ruolo di RSPP della Struttura. Titoli di studio: Laurea in Scienze Politiche e Diploma di Specializzazione in Diritto Amministrativo e Scienza dell'Amministrazione.



Gianluigi Deiana ian.deiana@inaf.it

Diploma di maturità scientifica e passione per l'astronomia fin da piccolo. Socio di 2 associazioni astrofili del territorio regionale. Ho iniziato a lavorare alla Stazione Astronomica di Cagliari a metà degli anni 80 e sono adesso Collaboratore Tecnico Enti di Ricerca. Mi occupo di sicurezza dell'Osservatorio di Cagliari e del sito del Sardinia Radio Telescope (SRT), e svolgo anche attività didattiche e divulgative. Gestisco i dati meteorologici del sito SRT per schedulare le osservazioni.

Nicola Di Cicco nicola.dicicco@inaf.it

Laureato magistrale in architettura è attualmente in servizio presso l'INAF-OAPd in qualità di responsabile dei servizi generali e di RSPP. Svolge funzioni di coordinamento di attività edilizie, tecnologiche e impiantistiche, finalizzate al mantenimento e all'adeguamento normativo-funzionale delle attività istituzionali di ricerca astronomica e scientifiche di laboratorio, nonché alla tutela del patrimonio storico e architettonico della sede di epoca medievale di INAF-OAPd.



Stefano Giovannini stefano.giovannini@inaf.it

Primo Tecnologo INAF, attuale responsabile dei Servizi Informatici gestionali e ammini-

strativi e della Transizione al Digitale, settore di cui mi occupo dal mio ingresso in INAF nel 2005. Dal 2015 faccio inoltre parte del Consiglio di Amministrazione dell'ente in qualità di membro eletto dalla comunità e coordino le attività e le analisi all'interno del cosiddetto Controllo di Gestione.

Davide Inamo davide.inamo@inaf.it

Dipendente dell'Osservatorio di Trieste dal 2006, oltre all'incarico di RSPP ricopre il ruolo di responsabile del servizio Edilizia ed Appalti. In precedenza ha svolto la libera professione di geometra e successivamente è stato dipendente in qualità di funzionario tecnico presso un Comune della provincia di Trieste quale responsabile della manutenzione e servizi esterni, dell'urbanistica e dell'ambiente.



Jacopo Lenzi jacopo.lenzi@inaf.it

Ho conseguito il diploma di geometra e abilitazione alla professione. Ho lavorato come libero professionista in uno studio tecnico occupandomi di sicurezza sui luoghi di lavoro, sicurezza antincendio ed edilizia in generale. Dal 2021 lavoro in INAF all'Osservatorio Astrofisico di Arcetri all'interno dei Servizi Generali, occupandomi della gestione dei servizi e delle manutenzioni delle strutture che compongono l'osservatorio. Dal settembre 2022 ricopro il ruolo di RSPP per la sede di Arcetri.

Giovanni Liggio giovanni.liggio@inaf.it

Geometra e laurea in Scienze Politiche e Relazioni Internazionali. Ha lavorato presso l'Ufficio Tecnico del Comune di Corleone dal 1988 al 1993. Dal 1993 lavora presso l'INAF-Osservatorio Astronomico di Palermo (CTER) con i ruoli di Responsabile dell'Ufficio Tecnico e Responsabile del Servizio di Prevenzione e Protezione. Si occupa del Building manager ed ha acquisito ampia esperienza sui Lavori Pubblici e acquisti di beni e servizi. Fa parte dello staff del servizio di meteorologia dell'OAPa.

Sergio Mariotti sergio.mariotti@inaf.it

Nascita: 1964, Diploma di maturità: 1983. Tecnico specializzato con una comprovata storia di lavoro nel settore della ricerca scientifica a partire dall'anno 1984. Esperto nei settori tecnologici quali criogenia e radiofrequenza. Dall'anno 2010 ricopre l'incarico di RSPP.

Francesco Massaro francesco.massaro@inaf.it

Diploma di Geometra e Laurea in Scienze della Prevenzione. Dopo una decennale esperienza nel privato, ho iniziato la mia attività all'Osservatorio Astronomico di Roma. Faccio parte della squadra addetta alla prevenzione incendi, di evacuazione e di primo soccorso con l'uso del DAE. Sono inoltre Responsabile dell'Ufficio tecnico e dei Servizi generali, Responsabile del Servizio di Prevenzione e Protezione, e Referente Nazionale per l'INAF della Banca Dati della Pubblica Amministrazione (BDAP).



Barbara Neri barbara.neri@inaf.it

Lavoro attualmente in INAF in particolare al SID (Sistemi Informativi per il Digitale) di diretta collaborazione con la Direzione Generale INAF: gestione e mantenimento dei gestionali amministrativi attualmente in dotazione in INAF (Epas rilevazione presenze, TEAM

software di contabilità, Civilia per acquisizione fatturazione elettronica, HR gestionale del personale, CSA Carriere e stipendi, Pica per il reclutamento). Laurea Magistrale in Economia. Master di primo livello in City management.



Giovanni Occhipinti giovanni.occhipinti@inaf.it

Lavoro presso l'Osservatorio Astrofisico di Catania come tecnico ottico-meccanico. Mi occupo della gestione della sicurezza sul lavoro in quanto RSPP. Collaboro con il gruppo di Astrofisica Sperimentale (LasP) e con il gruppo "Catania astrophysical Observatory of Detectors (COLD)". Come attività di ricerca mi occupo di rintracciare e verificare siti di interesse archeoastronomico.

Mario Pepe mario.pepe@inaf.it

Collaboratore Tecnico presso la sede di Merate dell'INAF Osservatorio Astronomico di Bre-
ra (OABr) dal 2008. Laureato in Scienze per l'Ambiente e il
Territorio, si occupa principalmente di sicurezza sul lavoro e di affidamento di contratti di
lavori, servizi e forniture. Nominato RSPP dal 2012 prima solo per la sede di Merate, poi
anche per la sede di Milano dell'OABr ed in ultimo per lo IASF di Milano, è stato tra i promo-
tori ed utilizzatore dei software di gestione della sicurezza in INAF.



Francesco Perrotta francesco.perrotta@inaf.it

Laureato in Ingegneria Aeronautica, ha lavorato dal 1997 nel campo della progettazione di te-
lescopi e strumentazione di piano focale (analisi strutturale, ottica attiva). Funzionario tecnico
all'Osservatorio Astronomico di Capodimonte (OACN) dal 2000, poi Tecnologo INAF dal 2007. At-
tualmente è Responsabile del Servizio Affari Generali e Tecnici e del Servizio di Prevenzione e Pro-
tezione dell'INAF- OACN. In ambito PNRR è impegnato nella gestione amministrativa degli appalti
in capo all'INAF-OACN.

Pier Raffaele Platania pier.platania@inaf.it

Pier Raffaele Platania è un Perito Industriale dell'Istituto Nazionale di Astrofisica, in forza
presso la sede osservativa del Radiotelescopio di Noto.
È Responsabile del Servizio di Prevenzione e Sicurezza sul Lavoro, dei servizi Generali
e gestione patrimonio, fornisce supporto alla Direzione per la risoluzione di situazioni a
carattere emergenziale con efficace applicazione di problem solving e decision making. Si oc-
cupa, inoltre, da anni di divulgazione scientifica e fotografia.



Irene Zago support@agews.atlassian.net

Laureata magistrale in Sicurezza Alimentare, dopo gli studi si avvicina al mondo HSE ed
alla consulenza. Attualmente lavora come Project Manager Senior per AGE Web Solutions,
realtà dinamica che si occupa della progettazione, sviluppo, manutenzione e vendita di
soluzioni software verticali e servizi per Aziende ed Enti Pubblici, dedicati a sicurezza negli
ambienti di lavoro, sistemi di gestione di qualità, ambiente, sostenibilità e sicurezza, e relativi
servizi di consulenza integrata HSE/ESG.



Phishing simulato: le campagne di phishing sono tutte equivalenti?

Fabrizio Fioravanti, Marius Bogdan Spinu, Marco Alamanni

Università degli Studi di Firenze

Abstract. Nel presente articolo, partendo dalla consapevolezza che il fattore umano è uno degli elementi più critici nella catena della sicurezza informatica, viene presentato l'utilizzo di campagne di phishing come strumento di valutazione dei reali rischi aziendali e quindi finalizzate all'aumento della consapevolezza del rischio cyber degli utenti all'interno dell'Organizzazione. L'analisi è stata quindi estesa alle tipologie di campagne di phishing simulato studiandone gli impatti sull'organizzazione

Keywords. cybersecurity, awareness, phishing, training, risk

Introduzione

La tematica della sicurezza informatica è diventata ultimamente sempre più rilevante sia per il significativo aumento della disponibilità di servizi online (quindi un numero maggiore di sistemi informatici target) sia per una situazione geopolitica particolarmente complessa a livello mondiale.

L'analisi della consapevolezza (awareness) in ambito di cybersecurity è un tema molto sentito a livello nazionale (piano triennale per l'informatica nella PA – CAPITOLO 6 [1]), ma ovviamente è comunque di interesse primario per le organizzazioni identificare a quali rischi l'organizzazione stessa sia più esposta in termini di sicurezza informatica.

Una delle principali tematiche da considerare nella valutazione della postura di sicurezza di una organizzazione rimane il fattore umano e come esso possa rendersi complice talvolta inconsapevole di gravi compromissioni di sicurezza che possano portare anche a data breach rilevanti; uno degli elementi principali da valutare in questo contesto è come reagisca l'organizzazione a fronte di tentativi di phishing, in quanto la compromissione dell'identità digitale a mezzo di campagne di phishing è un tema sempre più attuale, diffuso e monetizzabile in vari modi da gruppi hacker.

Phishing simulato come strumento di indagine

Le campagne di phishing simulato, ove si inviano false e-mail di phishing con lo scopo di verificare come il bersaglio di queste campagne reagisca, è uno strumento importante non tanto per puntare l'attenzione sui comportamenti dei singoli, ma bensì per capire a quali rischi l'organizzazione nel suo complesso sia più esposta, programmando quindi percorsi informativi e formativi maggiormente efficaci.

Nel 2020 è iniziata una ricerca di mercato per individuare potenziali strumenti capaci di inviare campagne di phishing efficaci e nel 2021 e 2022 sono stati adottati 2 prodotti

commerciali diversi per inviare a campioni casuali del personale dell'Ateneo campagne di phishing simulato.

La scelta dei prodotti, oltre che per la loro specifica usabilità è stata guidata anche dalla letteratura esistente in termini di campagne di phishing e della loro capacità di generare campagne con l'efficacia desiderata [2][3].

Di conseguenza, durante gli anni 2021 e 2022 all'interno dell'Ateneo Fiorentino sono state condotte analisi in questo senso, ma questo articolo non si concentra sui risultati numerici e sulle percentuali di persone che hanno avuto un certo comportamento bensì sugli strumenti adottati e sulle informazioni che sono state ricavate usandoli in maniera propria ed efficace per trarre informazioni sull'organizzazione nel suo complesso.

Le campagne di phishing non sono infatti tutte equivalenti e per avere una valutazione reale dei rischi a cui è esposta l'organizzazione è necessario non solo valutare la percentuale di persone che hanno fatto click su link malevoli o che hanno fornito le loro credenziali, ma anche e soprattutto gli stimoli a cui sono stati sottoposti.

Un aspetto rilevante considerato nella progettazione delle diverse campagne è stato quello di realizzarle a partire da informazioni e strumenti reperibili pubblicamente (elenchi di mail, pagine web, relazioni gerarchiche, etc) che fossero quindi anche nella disponibilità di un potenziale attaccante.

Il NIST (National Institute of Standards and Technology) ha proposto infatti di categorizzare una campagna di phishing sia sulla base del numero di indizi (number of cues) che sono presenti nella campagna stessa, ma anche su quanto la campagna sia correlata all'ambiente lavorativo o personale (premise alignment) del bersaglio.

Per indizi si intendono normalmente le caratteristiche quali, ad esempio, il fatto che ci siano link non pertinenti, mittenti esterni all'organizzazione oppure che la campagna voglia suggerire un senso di urgenza o che abbia errori di ortografia.

L'influenza predominante che anche nella nostra analisi abbiamo verificato essere presente, riguarda proprio questo ultimo aspetto, in quanto costruendo campagne con lo stesso numero e tipologia di indizi, la percentuale di persone che cadono nel phishing è decisamente più elevata se la simulazione riporta a contesti lavorativi o personali del soggetto.

Nella Tab. 1 è stata riportata anche una schematizzazione della difficoltà percepita dal soggetto nell'identificare come phishing la simulazione sulla base di questi fattori.

Tab. 1
Esempio di impatto dei diversi fattori di una campagna di phishing sulla complessità percepita

Number of cues	Premise alignment	Detection difficulty
Few (more difficult)	High	Very difficult
	Medium	Very difficult
	Low	Moderately difficult
Some	High	Very difficult
	Medium	Moderately difficult
	Low	Moderately to Least difficult
Many (less difficult)	High	Moderately difficult
	Medium	Moderately difficult
	Low	Least difficult

Nelle campagne erogate in Ateneo il numero di indizi era sempre basso, anche se alcuni elementi come link non consoni e mittenti non istituzionali erano sempre presenti proprio come spia di un phishing, al fine di permettere di rendere riconoscibile la campagna come tale.

Nella nostra esperienza infatti abbiamo verificato che sottoponendo a campagne di phishing generaliste il campione di soggetti, notavamo una buona postura di sicurezza generale dell'organizzazione mentre la risposta cambiava molto, a parità di difficoltà di identificazione della mail di phishing, se modificavamo la rilevanza a livello lavorativo o si applicava a contesti abituali per l'utente.

Quanto più il phishing era correlato all'attività lavorativa ed all'uso abituale dell'e-mail ed a strumenti quotidiani dell'attività lavorativa, tanto più si alzavano le percentuali di successo del phishing stesso.

Questa analisi ha permesso quindi di esaminare ed estrapolare con precisione la reale postura di sicurezza ed identificare a quali rischi le diverse categorie del personale fossero più esposte tracciando un profilo abbastanza preciso dell'utenza e degli ambiti di vulnerabilità; questo non sarebbe stato possibile utilizzando strumenti che sottoponevano il campione a campagne di phishing casuali senza una valutazione precisa sia della difficoltà di individuazione e dell'attinenza al soggetto obiettivo dell'analisi. La creazione quindi di campagne mirate controllando tutti gli aspetti dalla progettazione della campagna insieme alla resa grafica di essa ed al target ha permesso di ricavare importanti informazioni su come proteggere meglio l'organizzazione da questo rischio.

L'analisi si è spinta fino alla creazione di simulazioni di phishing molto mirato focalizzandosi sulle possibili leve a cui il soggetto poteva essere sensibile. In questo modo chi è caduto nel tranello della simulazione, ha permesso proprio di identificare le potenziali aree di criticità all'interno dell'organizzazione e non soltanto di fare statistiche circa le persone avevano avuto un certo tipo di risposta; questo insieme di informazioni ha fornito quindi indicazioni su quali fossero gli elementi su cui porre maggiore attenzione sia a livello organizzativo che in termini di proposizione di percorsi formativi ed informativi oltre che indicazioni puntuali a particolari categorie di utenti particolarmente rilevanti all'interno dell'organizzazione per il loro ruolo.

Conclusioni

Dopo 2 anni di sperimentazione e alcune campagne di phishing simulato erogate la nostra conclusione conferma l'ipotesi che per ottenere informazioni rilevanti sui rischi correlati al phishing non basti erogare campagne generaliste uguali per tutti o estratte casualmente da un pool di campagne non specializzate, soprattutto se tali campagne non sono state classificate per difficoltà e per attinenza, mentre la progettazione di campagne mirate permette di valutare al meglio la reale postura di sicurezza dell'organizzazione.

Le campagne di phishing non sono quindi tutte equivalenti e la realizzazione di una efficace analisi del rischio phishing passa da una progettazione mirata e specifica delle campagne a cui sottoporre il campione selezionato oltre che dalla selezione di informazioni pubblicamente accessibili, più che dall'erogazione massiva di simulazioni generaliste.

Questa analisi conferma e rende evidente il ruolo del “social engineering” e la rilevanza che possono assumere le informazioni che un ente pubblico è vincolato a rendere disponibili pubblicamente (e di conseguenza anche ai potenziali attaccanti) e che possono quindi contribuire a dare maggiore efficacia ad un reale attacco phishing sfruttando il fattore umano.

Riferimenti bibliografici

- [1] Piano Triennale per l'informatica nella PA – Aggiornamento 2022/2024 - https://www.agid.gov.it/sites/default/files/repository_files/piano_triennale_per_linformatica_nella_pa_2022-2024.pdf
- [2] Michelle Steves, Kristen Greene and Mary Theofanos - A Phish Scale: Rating Human Phishing Message Detection Difficulty - Workshop on Usable Security (USEC) 2019
- [3] Michelle Steves, Kristen Greene and Mary Theofanos - Categorizing human phishing difficulty: a Phish Scale - Journal of Cybersecurity, 2020, 1–16

Autori



Fabrizio Fioravanti fabrizio.fioravanti@unifi.it

Fabrizio Fioravanti è attualmente responsabile delle Unità di Processo Sistemi, tecnologie cloud e di sicurezza informatica e ad interim di Reti e postazioni di lavoro. Laureato in Ingegneria Elettronica ha un dottorato di ricerca in Ingegneria Informatica e delle Telecomunicazioni e da oltre 25 anni lavora nel settore ICT. Ha diverse pubblicazioni scientifiche al suo attivo in riviste italiane ed internazionali oltre che in libri e conferenze.

Marius Bogdan Spinu marius.spinu@unifi.it

Ingegnere, Dirigente dell'Area per l'Innovazione e Gestione dei Sistemi Informativi ed Informatici nonché Responsabile per la Transizione al Digitale dell'Università degli Studi di Firenze. Lunga esperienza come dirigente nel Sistema Sanitario Toscano nell'ambito ICT con riferimento ai processi ospedalieri, amministrativi e della logistica del farmaco. Crede che la tecnologia sia uno dei principali driver dei processi di crescita e che la sicurezza sia un elemento trasversale di tali processi.



Marco Alamanni marco.alamanni@unifi.it

Marco Alamanni si occupa attualmente di sicurezza informatica, all'interno dell'Unità di processo "Sistemi, tecnologie cloud e di sicurezza informatica" del Sistema Informatico dell'Ateneo.

È laureato in Informatica presso l'Università di Pisa, dove ha anche frequentato il corso di Laurea Magistrale in Sicurezza Informatica.

Tra i suoi interessi principali nell'ambito della sicurezza informatica, ci sono threat intelligence, digital forensics e Osint.

Cyber Resilience Act and Open Source: problemi aperti e prospettive future

Nadina Foggetti

Consiglio Nazionale delle Ricerche IBIOM - Bari

Abstract. L'Unione Europea ha recentemente innovato la normativa in materia di cyber security, introducendo diverse norme volte a definire un nuovo quadro giuridico di riferimento per gli Stati membri dell'UE in questo settore. In particolare il Regolamento Il Cyber Resilience Act (RCA) si inserisce nel contesto della trasformazione digitale avviata dall'Europa e da attuarsi entro il 2030, ponendosi in maniera complementare al Regolamento 2019/881 relativo all'ENISA, attraverso il quale è stato introdotto nell'ordinamento comunitario un quadro comune di certificazione della cybersicurezza per le tecnologie dell'informazione e della comunicazione ("TIC"). Obiettivo principale del Regolamento è quello di definire degli standard comuni di sicurezza informatica per i prodotti digitali connessi in rete (cd. "IoT") e per i relativi servizi al fine di proteggere i consumatori e il mercato dagli incidenti informatici, salvaguardando le imprese e gli utenti che acquistano o utilizzano prodotti, o software, con componenti digitali. All'interno dell'ambito di applicazione della nuova normativa rientrano anche i prodotti e alle tecnologie open source soprattutto all'interno di Servizi ICT a supporto delle infrastrutture di ricerca. In questo lavoro si analizzeranno le principali novità introdotte della nuova disciplina europea ed il loro impatto con la tutela dei principi fondamentali, quali la libera circolazione nel mercato dei prodotti digitali, i principi dell'open Science e dell'Open Source

1. Il nuovo quadro giuridico di riferimento

All'interno della legislazione europea nel contesto della cyber security, nel 2013 è entrata in vigore la direttiva 2013/40/UE relativa agli attacchi contro i sistemi di informazione, che armonizza la criminalizzazione e le sanzioni per i reati contro i sistemi di informazione. In seguito è stata adottata la direttiva (UE) 2016/1148 sulla sicurezza delle reti e dei sistemi informativi (direttiva NIS), primo strumento legislativo europeo sulla cybersicurezza. La proposta di Regolamento relativo ai requisiti orizzontali di cybersicurezza per i prodotti con elementi digitali e che modifica il regolamento (UE) 2019/1020, il Cyber Resilience Act (CRA), e che si inserisce nel contesto della trasformazione digitale avviata dall'Europa ha permesso la definizione, nell'ordinamento europeo, di un quadro comune di certificazione della cybersicurezza per le tecnologie dell'informazione e della comunicazione ("TIC"). La nuova disciplina mira a garantire il miglioramento della sicurezza dei prodotti con elementi digitali fin dalla fase di progettazione e sviluppo e durante l'intero ciclo di vita, mediante la definizione di un quadro coerente in materia di cybersicurezza che faciliti la compliance per i produttori di hardware e software. Pone, inoltre, le basi per migliorare la trasparenza delle proprietà di sicurezza dei prodotti con elementi digitali e consentire alle imprese e ai consumatori di utilizzarli in modo sicuro.

Al fine di raggiungere questi obiettivi la nuova disciplina impone degli obblighi per le a-

ziende che producono strumenti digitali, o che li acquisiscono mediante supply chain.

2. Ambito di applicazione della disciplina

In merito all'applicazione *ratione materiae*, il Considerando 11 del CRA stabilisce che lo stesso “non dovrebbe disciplinare il software libero e open source sviluppato o fornito al di fuori di un'attività commerciale”. Nell'ambito del software Open Source (OSS), spesso l'accesso al servizio di assistenza, la fornitura di una piattaforma software o l'uso di dati personali contenuti in database è regolato dal pagamento di un prezzo.

Sotto il profilo dell'applicazione *ratione personae*, il Regolamento stabilisce che può essere considerato “fabbricante” qualsiasi persona fisica o giuridica che sviluppi o faccia progettare prodotti e li commercializzi con il proprio nome o marchio, a titolo oneroso o gratuito. A questa particolare categoria sono attribuiti obblighi quali la redazione di una valutazione dei rischi di cybersecurity, nonché l'impegno a gestire le vulnerabilità e ad adottare le misure di sicurezza correttive (ivi compreso l'eventuale ritiro del prodotto dal mercato) e l'onere di procedere all'informazione e alla comunicazione all'autorità competente. Similari oneri ricadono, ai sensi del CRA sui distributori di prodotti digitali. Rientra in questa categoria qualsiasi persona fisica o giuridica nella catena di approvvigionamento, diversa dal fabbricante o dall'importatore, che mette a disposizione un prodotto con elementi digitali sul mercato dell'Unione senza modificarne le proprietà. In particolare i distributori sono tenuti ad uno specifico obbligo di verifica rispetto alla presenza del marchio UE sulla cyber security.

L'articolo 3 comma 18, inoltre, prevede l'applicazione degli obblighi a coloro che producono software a titolo gratuito e l'articolo 16 del Regolamento, rubricato “Altri casi in cui si applicano gli obblighi dei fabbricanti”, ne estende la portata a qualsiasi persona fisica o giuridica che apporta una sostanziale modifica al prodotto con elementi digitali. In questa prospettiva qualsiasi programmatore, che contribuisce allo sviluppo di un software il cui uso è regolato da licenze libere diventa, giuridicamente, un produttore.

Inoltre, tra gli obblighi dei fabbricanti, vi è quello di segnalare al soggetto che si occupa della manutenzione, una vulnerabilità anche quando si tratta di un componente OS integrato nel prodotto con elementi digitali.

Sulla base di quanto emerge dobbiamo concludere che la disciplina trovi applicazione anche nei confronti del software e della tecnologia OS.

Se si amplia lo studio del legal framework alla proposta di direttiva sulla responsabilità per danno da prodotti difettosi (2022/0302(COD)), è possibile notare che l'articolo 4 stabilisce che per “prodotto” si intende anche il software; e che “fabbricante” identifica chi sviluppa, produce o fabbrica un prodotto per uso proprio, allineando di fatto l'ambito di applicazione a quello definito nel CRA.

La questione non è meramente tuzioristica, poiché influisce in modo significativo, non solo sugli obblighi del produttore (sviluppatore), ma anche sulle eventuali responsabilità in cui potrebbe incorrere in caso di mancata compliance con i requisiti prescritti.

3. L'applicazione alla tecnologia open source del CRA

Un primo aspetto rilevante, in termini di ricadute sull'OSS è quello che attiene alla di-

stinzione, operata all'interno del Regolamento, tra software finito e software non ancora rilasciato (versione beta). Il considerando 21 del Regolamento, stabilisce che è possibile condividere la versione beta a condizione che sia messa a disposizione solo per il tempo necessario a testarla e a raccogliere riscontri. In questa ipotesi, i fabbricanti dovrebbero provvedere affinché il software messo a disposizione a tali condizioni sia rilasciato solo a seguito di una valutazione dei rischi e sia conforme ai requisiti di sicurezza relativi alle proprietà dei prodotti con elementi digitali imposti dal CRA.

Una delle caratteristiche dell'OSS è l'innovazione, che si concretizza nella possibilità di garantire una grande flessibilità e libertà nel poter cambiare elementi senza eccessiva restrizione. In ambiente OSS è difficile distinguere tra versioni Beta e versioni definitive, proprio perché il software è in continua evoluzione ed in continuo cambiamento, per questo è qualificazione dello status dello stesso, ai fini dell'applicazione delle norme in materia di cybersecurity, è particolarmente complessa. Partendo dall'assunto che l'OSS è, altresì in continua evoluzione, il rispetto degli obblighi di documentazione, diverrebbe particolarmente oneroso, anche nella prospettiva del continuo aggiornamento della stessa rispetto alle diverse versioni.

Una caratteristica che contraddistingue l'OSS è la continua produzione di lavoro creativo sui software stessi, operato dalla community di sviluppatori. Questo rende particolarmente difficile tenere traccia degli apporti creativi e pone l'annoso problema, più volte discusso della dottrina, della disciplina giuridica da applicare alle parti aggiunte in seguito al primo sviluppo del software. Si discute se debba essere qualificato come opera collaborativa, creata da più autori, oppure se il software sviluppato di volta in volta rappresenti un'opera derivata da quella originale. Quando si opera in un contesto transnazionale, quale è quello della ricerca scientifica e tecnologica, occorre tenere conto delle differenze che ci sono tra gli ordinamenti giuridici che compongono il legal framework di riferimento. In particolare, all'interno dei sistemi di common law, l'OSS si qualifica come opere derivata, mentre negli ordinamenti di civil law, si parla di elaborazioni creative. La distinzione è necessaria ai fini dell'attribuzione dei diritti morali d'autore sull'opera, ma anche per individuare la titolarità degli obblighi previsti dal Regolamento.

Inoltre, l'applicazione di obblighi e responsabilità nei confronti dello sviluppatore del software è in grado di determinare un conflitto con le disposizioni contrattuali inserite all'interno della licenza d'uso dell'OSS.

Queste ultime contengono espressi esoneri di responsabilità, in ragione del fatto che gli sviluppatori non percepiscono un compenso per il lavoro di sviluppo. Questa disposizione contrattuale, non esclude in nessun caso la responsabilità per dolo o colpa grave, ma potrebbe non comprendere quella per prodotti difettosi come richiesto dalla nuova disciplina in corso di approvazione.

L'applicazione del nuovo quadro giuridico europeo in materia di cyber security avrebbe, sulla base di quanto evidenziato, un impatto significativo sull'ecosistema open source. I soggetti destinatari degli obblighi e delle relative sanzioni previste dal CRA comprendono a pieno titolo gli sviluppatori di software open source, sebbene l'attività svolta dagli stessi sia a titolo gratuito. L'applicazione delle norme in parola è in grado di ridurre

in modo determinante il numero dei potenziali sviluppatori di software OS, in ragione delle importanti sanzioni eventualmente applicabili in caso di violazione. La circostanza, infatti, che gli sviluppatori contribuiscano con il proprio codice, permette l'inserimento degli stessi tra le categorie a cui la nuova disciplina europea attribuisce obblighi specifici e responsabilità.

La produzione di software OS potrebbe subire una significativa riduzione, in ragione degli eccessivi costi di compliance che risulterebbero essere gravosi e demotivanti per i singoli sviluppatori che intendono contribuire a progetti OS.

In secondo luogo vi sarebbe una limitazione alla circolazione del software open source a livello internazionale, dato il legal gap definito dalla differenza di regime giuridico tra le parti di codice che sono sviluppate ed utilizzate all'interno dell'UE e quelle sviluppate al di fuori dello spazio giuridico europeo. Questo gap potrà avere delle ricadute negative sulle dinamiche di interazione tra sviluppatori provenienti da diversi Paesi.

A fronte di queste criticità di impatto della normativa, le principali associazioni di software libero sono intervenute con una lettera aperta rivolta alle istituzioni dell'Unione Europea al fine di poter contribuire al dialogo teso all'elaborazione del testo definitivo del CRA. L'obiettivo è quello di garantire che le caratteristiche uniche del software open source siano considerate a livello normativo e che il Cyber Resilience Act non danneggi involontariamente l'ecosistema open source. Nella lettera si auspica l'apertura di un canale di dialogo, durante il processo di legislazione, al fine di assicurare che qualsiasi evoluzione dettata dal CRA, tenga conto della diversità delle pratiche di sviluppo di software open source, aperte e trasparenti, e che lo stesso divenga un meccanismo istituzionale di collaborazione fra le istituzioni europee e la community open-source, per garantire che le future decisioni legislative e politiche siano fondate su informazioni corrette. (https://elettronica-plus.it/lettera-aperta-per-il-cyber-resilience-act_128149/)

Bibliografia

Direttiva 2013/40/UE del Parlamento europeo e del Consiglio, del 12 agosto 2013, relativa agli attacchi contro i sistemi di informazione e che sostituisce la decisione quadro 2005/222/GAI del Consiglio, GU L 218 del 14.8.2013, pag. 8

Direttiva (UE) 2016/1148 del Parlamento europeo e del Consiglio, del 6 luglio 2016, recante misure per un livello comune elevato di sicurezza delle reti e dei sistemi informativi nell'Unione

GU L 194 del 19.7.2016, pag. 1.

Direttiva (UE) 2022/2555 del Parlamento europeo e del Consiglio del 14 dicembre 2022 relativa a misure per un livello comune elevato di cibersecurity nell'Unione, recante modifica del regolamento (UE) n. 910/2014 e della direttiva (UE) 2018/1972 e che abroga la direttiva (UE) 2016/1148 (direttiva NIS 2), GU L 333, del 27.12.2022, pag. 80.

Regolamento (UE) 2019/881 del Parlamento europeo e del Consiglio, del 17 aprile 2019, relativo all'ENISA, l'Agenzia dell'Unione europea per la cibersecurity, e alla certificazione della cibersecurity per le tecnologie dell'informazione e della comunicazione, e che abroga il regolamento (UE) n. 526/2013 ("regolamento sulla cibersecurity"), GU L 151 del

7.6.2019, pag. 15.

Proposta di Regolamento del Parlamento europeo e del Consiglio relativo a requisiti orizzontali di cibersecurity per i prodotti con elementi digitali e che modifica il regolamento (UE) 2019/1020, COM/2022/454 final, disponibile alla seguente URL: <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:52022PC0454#footnote4> (ultima consultazione 25 settembre 2023).

Proposta di direttiva on liability for defective products, Procedure 2022/0302/COD, disponibile alla seguente URL: https://eur-lex.europa.eu/procedure/EN/2022_302 (ultima consultazione 25 settembre 2023).

Dell'aversana F., Quale copyright per il mondo del file sharing? Il caso dei social network, in *Informatica e diritto*, XLIII annata, Vol. XXVI, 2017, n. 1-2, pp. 519-538.

Heim T N, Wessel R A, The Global Regulation of Cybersecurity: A Fragmentation of Actors, Definitions and Norms, in Lucía Millán Moro and Gloria Fernández Arribas (eds.), *Ciberataques y Ciberseguridad en la Escena Internacional*, 2020.

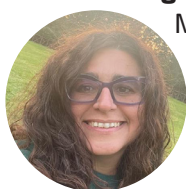
Kasper A, Antonov A, Towards Conceptualizing EU Cybersecurity Law, *ZEI*, 2019.

Markopoulou D, Papakonstantinou V, De Hert P, "The new EU cybersecurity framework: The NIS Directive, ENISA's role and the General Data Protection Regulation", *Computer Law and Security Review*, Vol. 35 Issue 6, November 2019.

McGowan D., Legal Implications of Open-Source Software, Symposium: Intellectual Property Challenges in the Next Century, in *University of Illinois Law Review*, Vol. 2001, pp. 241-304.

Papakonstantinou V., Cybersecurity as praxis and as a state: The EU law path towards acknowledgement of a new right to cybersecurity?, in *Computer Law & Security Review* Volume 44, 2022, p. 44 ss.

Biografia



Nadina Foggetti

Tecnologo CNR – IBIOM Bari, già Assegnista di ricerca Senior INFN Bari. Laurea in Giurisprudenza, Master in diritto comunitario e transnazionale -UniTN, Corso avanzato in Data Protection e Data Governance presso e in Coding per Avvocati e Legal Tech -Unimi. Dottore di ricerca in diritto internazionale e dell'Unione europea presso l'Università degli studi di Bari A. Moro. Avvocato e mediatore. Membro SIDI. dal 2005 collabora ad importanti progetti di ricerca nazionali e internazionali nei settori del cybercrime, cybersecurity, data privacy, diritto informatico e cloud computing. Docente di diritto internazionale e diritto della cybersecurity Corso di Laurea in Scienze della Mediazione Linguistica. Autore di diverse pubblicazioni nel settore. Attualmente collabora al progetto ELIXIRxNextGenIT sui temi dell'Open Science, Fair, Open Access.

BullyBuster - A framework for bullying and cyberbullying action detection by computer vision and artificial intelligence methods and algorithms

Wanda Nocerino¹, Vincenzo Gattulli², Stefano Marrone³, Giulia Orrù⁴, Grazia Terrone⁵, Antonio Galli³, Michela Gravina³, Donatella Curtotti¹, Donato Impedovo¹, Gian Luca Marcialis⁴, Carlo Sansone³

¹Università di Foggia, ²Università degli Studi di Bari "Aldo Moro", ³Università degli Studi di Napoli "Federico II", ⁴Università di Cagliari, ⁵Università degli Studi di Roma "Tor Vergata"

Abstract. Le azioni di bullismo e cyberbullismo, connotate da comportamenti violenti e di prepotenza, reiterati nel tempo e condotti a danno di bambini o adolescenti, sono diventate una drammatica emergenza. Nonostante gli sforzi profusi a livello legislativo, sono ancora embrionali gli studi volti a proteggere l'individuo e il suo benessere in un momento antecedente alla verifica dell'evento attraverso la prevenzione delle azioni di bullismo e di cyberbullismo. Nel 2017, nell'ambito del bando PRIN, il Ministero dell'Istruzione e della Ricerca (MIUR), ha finanziato il progetto dal titolo "BullyBuster - A framework for bullying and cyberbullying action detection by computer vision and artificial intelligence methods and algorithms", presentato da quattro RU delle Accademie del sud Italia (Napoli "Federico II", Bari, Cagliari, Foggia), con l'obiettivo di offrire un supporto concreto per la prevenzione e il contrasto del fenomeno. Nel 2023, BullyBuster è stato incluso nella top-100 della "lista globale di progetti che risolvono problemi legati ai 17 Obiettivi di Sviluppo Sostenibile delle Nazioni Unite dall'associazione IRCAI-UNESCO"

Keywords. Bullying, deepfake detection, text analysis, keystroke dynamics

1. Lo stato dell'arte

Le azioni di bullismo e cyberbullismo, connotate da comportamenti violenti e di prepotenza, ripetuti e continuati nel tempo, fisici o a mezzo internet, sono diventate una drammatica emergenza tra gli adolescenti.

Uno studio realizzato da HBSC nel 2018 (https://www.salute.gov.it/portale/temi/p2_6.jsp?lingua=italiano&id=2954&area=stiliVita&menu=sorve), registra un aumento dei casi del 70 % tra i minori in età compresa tra i 14 e i 16 anni e anche tra gli adulti (40%). A seguito della pandemia da Covid-19, il fenomeno è divenuto ancora più allarmante: la ONG internazionale "Bullismo senza frontiere" ha condotto una ricerca tra gennaio 2021 e febbraio 2022 (<https://www.paidea.it>), evidenziando come tali fenomeni e, in particolare, quelli di cyberbullismo, sono aumentati in ragione del crescente tempo condiviso online attraverso un utilizzo pressoché continuo, nell'arco della giornata, delle piattaforme come Instagram, TikTok, Teams e Zoom.

Diversi studi emergenti hanno dimostrato che il bullismo e il cyberbullismo possono avere un grave impatto fisico e psicologico, ad esempio sintomi psicosomatici e depressivi, ansia, comportamento autolesionista e abuso di sostanze (Goebert D. et al, 2011). Pertanto, sono necessari programmi di prevenzione e contrasto per neutralizzare le conseguenze e gli effetti di questi fenomeni.

Nonostante l'entrata in vigore della l. 71/2017, che ha introdotto forme di tutela per il contrasto del bullismo e del cyberbullismo, allo stato dell'arte si riscontrano criticità in ragione dell'assenza di una previsione normativa unitaria ed organica che, al contrario, appare ancora oggetto di una disciplina frammentaria.

Ciò comporta numerose difficoltà interpretative, tanto che la giurisprudenza è costretta a sussumere le condotte di bullismo nell'alveo di singole fattispecie già esistenti nell'ordinamento (es.: minaccia, stalking, percosse, lesioni, furto di identità digitale, ecc.).

Al fine di colmare la lacuna normativa con precipuo riferimento al cyberbullismo, il legislatore (l. 69/2019) ha tipizzato la fattispecie di diffusione illecita di immagini o video sessualmente espliciti (c.d. revenge porn). Anche in questo caso, tuttavia, la disciplina non appare del tutto soddisfacente, dal momento che non considera sufficientemente gli aspetti legati alla prevenzione (rectius: neutralizzazione) del fenomeno, agendo solo a posteriori, mediante un'istanza di oscuramento o attraverso la rimozione o il blocco di qualsiasi altro dato personale della vittima.

Parallelamente agli sforzi profusi dal legislatore, sono diverse le ricerche scientifiche che si occupano dello studio del fenomeno (Barlett, C.P. et al, 2021; Das S. et al, 2020; Mkhize S. e Gopal N., 2021) nella prospettiva di garantire un supporto alle vittime di bullismo e cyberbullismo. Deve, tuttavia, evidenziarsi come siano ancora embrionali gli studi volti a proteggere l'individuo e il suo benessere in un momento antecedente alla verifica dell'evento, cioè, attraverso la prevenzione delle azioni di bullismo e di cyberbullismo.

2. Il progetto "BullyBuster"

In questo contesto, nel 2019, nell'ambito del bando relativo ai Progetti di Rilevante Interesse Nazionale (PRIN), il Ministero dell'Istruzione e della Ricerca (MIUR), ha finanziato l'ambizioso progetto dal titolo "BullyBuster - A framework for bullying and cyberbullying action detection by computer vision and artificial intelligence methods and algorithms", presentato da quattro RU delle Accademie del sud Italia (Napoli "Federico II", Bari, Cagliari, Foggia). Nel 2023, BullyBuster è stato incluso nella top-100 della "lista globale di progetti che risolvono problemi legati ai 17 Obiettivi di Sviluppo Sostenibile delle Nazioni Unite dall'associazione IRCAI-UNESCO".



Fig. 1

Logo of the project "BullyBuster—A framework for bullying and cyberbullying action detection by computer vision and artificial intelligence methods and algorithms" (a). The project has been included in the Global Top 100 list of AI projects by IRCAI (b).

Si tratta di una proposta che - sfruttando un approccio integrato interdisciplinare di ingegneria informatica, diritto e psicologia - mira alla prevenzione e al contrasto delle azioni di bullismo e cyberbullismo.

Più concretamente, in questo progetto i diversi Atenei stanno cooperando per la creazione e lo sviluppo di un software che, ricorrendo all'intelligenza artificiale (IA) e a tecniche di computer vision, consente di individuare e segnalare comportamenti violenti commessi da minori in danno dei coetanei.

Tale programma opera servendosi di diverse tipologie di dati ricavati mediante specifiche tecniche:

- 1) analisi basata su video, per la segmentazione della scena mediante descrittori strutturali temporali e spaziali, al fine di rilevare specifiche azioni di bullismo sulla base dei movimenti della folla intorno alla vittima e, ove possibile, alla sua espressione facciale (Orrù et al, 2021);
- 2) analisi testuale, mediante il rilevamento di parole e frasi tipiche di molestie informatiche, oppressione e stalking (Marrone, S. e Sansone, C., 2022);
- 3) analisi comportamentale mediante il rilevamento della dinamica della sequenza di tasti e l'emergente campo dell'analisi tattile (Gattulli et al, 2022).

Al fine di individuare correttamente gli indicatori (rectius: comportamenti) sintomatici delle azioni di bullismo e cyberbullismo, il programma viene tarato sulla base delle indicazioni provenienti da studi giuridici e psicologici, indispensabili per garantire l'effettività e l'efficacia dei risultati ottenuti. Di qui, sullo sfondo del quadro legislativo di riferimento, il software viene predisposto considerando i protocolli e i modelli psicologici generativi di azioni tipiche di bullismo e cyberbullismo.

Dopo aver strutturato il programma sulla base delle indicazioni giuridiche e psicologiche, i diversi tools sono destinati a trovare impiego in un'App per cellulari e computer denominata "BullyBuster" (Acchiappabulli) da installare - volontariamente e solo una volta che i genitori dei minori coinvolti abbiano prestato il consenso informato - sui dispositivi elettronici, da utilizzare a scopo sia repressivo che preventivo con sistemi di rilevamento dello stato emotivo dell'utente e di azioni o contenuti pericolosi intelligente (sistemi di videosorveglianza frasi minacciose o immagini o video contraffatti come deep fake) (Concas et al, 2022).

Un'indagine di questo tipo è funzionale non solo sul piano teorico per la costruzione di algoritmi efficaci per l'individuazione di comportamenti sintomatici di bullismo e cyberbullismo, ma anche sotto il profilo operativo: nell'ottica di favorire strategie e politiche di sviluppo delle infrastrutture digitali, lo strumento potrebbe trovare concreta attuazione per facilitare le azioni investigative e giudiziaria attraverso l'impiego degli elementi probatori raccolti sulla scena del crimine reale o virtuale.

Riferimenti bibliografici

Barlett, C.P., Rinker, A., Roth, B. (2021). Cyberbullying perpetration in the COVID-19 era: An application of general strain theory. *The Journal of Social Psychology*, 7, pp. 78-94;
Concas, S., La Cava, S.M., Orrù, G., Cuccu, C., Gao, J. (2022); Feng, X.; Marcialis, G.L.; Roli,

- F. Analysis of Score-Level Fusion Rules for Deepfake Detection. *Appl. Sci.* 12, p. 7365;
- Das, S., Kim, A., Karmakar, S. (2020). Change-Point Analysis of Cyberbullying-Related Twitter Discussions During COVID-19. *International Journal of Cyber Behavior, Psychology and Learning*, 5(4), pp. 31-52;
- Gattulli, V., Impedovo, D., Pirlo, G., Sarcinella, L. (2022), Cyber Aggression and Cyberbullying Identification on Social Networks. In *Proceedings of the 11th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2022)*, Vienna, Austria, pp. 644–651;
- Goebert D., Else I., Matsu C., Chung-Do J, Chang JY. (2011). The impact of cyberbullying on substance use and mental health in a multiethnic sample. *Maternal Child Health J*, 15(8), pp. 1282-1286;
- Marrone, S., Sansone, C. (2022), Identifying Users' Emotional States through Keystroke Dynamics. In *Proceedings of the 3rd International Conference on Deep Learning Theory and Applications, DeLTA, INSTICC, Lisbon, Portugal, 12–14 July 2022*; SciTePress: Setubal, Portugal, Volume 1, pp. 207–214;
- Mkhize, S., Gopal, N. (2021). Cyberbullying Perpetration: Children and Youth at Risk of Victimization during Covid-19 Lockdown. *International Journal of Criminology and Sociology*, 10, pp. 525-537;
- Orrù, G., Ghiani, D., Pintor, M., Marcialis, G.L., Roli, F. (2021), Detecting anomalies from video-sequences: A novel descriptor. In *Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR)*, Milan, Italy, pp. 4642–4649.

Autori

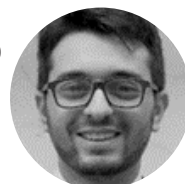


Wanda Nocerino wanda.nocerino@unifg.it

È dottore di ricerca in Scienze Giuridiche. Nel 2022 ha conseguito l'abilitazione scientifica nazionale alla funzione di professore di II fascia. Attualmente è RTDA (S.S.D. IUS/16), presso il Dipartimento di Giurisprudenza dell'Università di Foggia. È autrice di saggi, articoli, note a sentenza pubblicate sulle più autorevoli riviste nazionali ed internazionali e di due monografie. I suoi temi di ricerca includono le indagini digitali e le investigazioni preventive.

Vincenzo Gattulli vincenzo.gattulli@uniba.it

Laurea magistrale in Computer Science presso l'Università degli Studi di Bari Aldo Moro (UNIBA). Attualmente è Dottorando presso la stessa Università sulla tematica "Biometria Comportamentale". Ha collaborato a progetti relativi alla Cyber Security e al Machine Learning con pubblicazioni consecutive. In passato ha ricoperto il ruolo di consulente informatico presso lo spin-off accademico dell'UNIBA "Digital innoVAtion (DIVA)"



Stefano Marrone stefano.marrone@unina.it

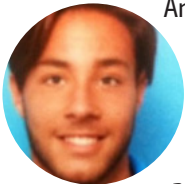
È ricercatore presso il Dipartimento di Ingegneria Elettrica e Informatica dell'Università di Napoli Federico II. I suoi temi di ricerca includono il pattern recognition e la computer vision, con applicazioni che vanno dall'elaborazione di immagini biomediche alle biometrie e alla guida autonoma. Più recentemente si è concentrato sull'etica e l'equità dell'IA in settori critici. È autore di oltre 50 pubblicazioni su riviste e atti di conferenze di rilevanza

internazionale.



Giulia Orrù giulia.orrù@unica.it

Ha conseguito il Dottorato di Ricerca in Ingegneria Elettrica e Informatica nel 2021. Attualmente è RTDA (Sistemi di Elaborazione delle Informazioni) presso l'Università degli Studi di Cagliari e afferisce al Dip. di Ingegneria Elettrica ed Elettronica. Dal 2014 fa parte dell'Unità Biometria del Pattern Recognition and Applications Lab e si occupa di elaborazione ed analisi di immagini e video nel campo delle tecnologie biometriche (impronte digitali, volti, deepfake, crowd analysis, etc.).



Antonio Galli antonio.galli@unina.it

È post-doc presso il Dipartimento di Ingegneria Elettrica e Informatica dell'Università degli Studi di Napoli Federico II. Ha conseguito la laurea triennale (2016) e magistrale con lode (2019) in Informatica e Ingegneria presso l'Università degli Studi di Napoli Federico II. I suoi principali interessi di ricerca sono nell'ambito del Deep Learning e della Big Data analytics con applicazioni che vanno dall'elaborazione di immagini ad elaborazione di serie temporali.



Donatella Curtotti donatella.curtotti@unifg.it

È professore ordinario di Diritto processuale penale presso il Dipartimento di Giurisprudenza dell'Università di Foggia nonché Direttore del medesimo Dipartimento. È autrice di oltre cento saggi, articoli, note a sentenza pubblicate sulle più autorevoli riviste nazionali ed internazionali e di tre monografie. È coordinatore del Corso di Laurea in Scienze Investigative. I suoi temi di ricerca includono le indagini tecnico-scientifiche e la sicurezza dello Stato.



Grazia Terrone grazia.terrone@uniroma2.it

È professore associato di Psicologia Dinamica presso Università di Roma Tor Vergata. È autrice di pubblicazioni nazionali e internazionali sull'analisi e la valutazione dei comportamenti esternalizzanti degli adolescenti e dei legami protettivi e di rischio con uso delle tecnologie. Membro e responsabile del ambito psicologico del progetto "BullyBuster" incluso nella top-100 lista globale di progetti che risolvono problemi legati ai 17 Obiettivi di Sviluppo Sostenibile delle Nazioni Unite dall'associazione IRCAI-UNESCO.



Michela Gravina michela.gravina@unina.it

È ricercatrice post-dottorato presso il Dipartimento di Ingegneria Elettrica e Tecnologie dell'Informazione dell'Università di Napoli Federico II. Ha conseguito la laurea magistrale in Ingegneria informatica e il dottorato di ricerca in Information and Communication Technology for Health. Le sue attività di ricerca si concentrano principalmente su tematiche di pattern recognition e computer vision, relativamente all'analisi di immagini mediche che includano diverse sorgenti di dati.



Donato Impedovo donato.impedovo@uniba.it

È professore associato presso il Dipartimento di Informatica dell'Università di Bari (IT). Ha gestito più di 35 progetti finanziati da istituzioni pubbliche e da PMI private. Ricopre il ruolo di revisore e relatore per l'UE nella valutazione dei progetti. I suoi attuali interessi di ricerca includono l'elaborazione dei segnali, pattern recognition, machine learning e la biometria. È coautore di oltre 140 articoli in questi settori sia su riviste internazionali che su atti di convegni. È socio IAPR.

Gian Luca Marcialis marcialis@unica.it

È professore associato presso l'Università di Cagliari e responsabile dell'Unità di Biometria. È fondatore e organizzatore delle otto edizioni della competizione internazionale Fingerprint Liveness Detection. È Principal Investigator e co-investigatore di progetti premiati come "BullyBuster", incluso nella top-100 lista globale di progetti che risolvono problemi legati ai 17 Obiettivi di Sviluppo Sostenibile delle Nazioni Unite dall'associazione IRCAI-UNESCO. Il Prof. Marcialis è membro senior dell'IEEE e membro dell'IAPR.



Carlo Sansone carlo.sansone@unina.it

È professore ordinario di Informatica presso il Dipartimento di Ingegneria Elettrica e Informatica dell'Università degli Studi di Napoli "Federico II". È autore di oltre 250 articoli su riviste internazionali e atti di convegni. I suoi interessi di ricerca coprono le aree dell'analisi e del riconoscimento delle immagini, del machine learning e del deep learning. Il Prof. Sansone ha coordinato diversi progetti nei settori dell'interpretazione di immagini biomediche, del rilevamento delle intrusioni di rete e del rilevamento del cyberbullismo utilizzando tecniche di intelligenza artificiale. È Fellow della IAPR e dell'AAIA.

Decidere al Ritmo degli Eventi

Massimo Carboni

GARR

Abstract. L'aumento esponenziale della potenzialità della nuova rete GARR-T comporta la necessità di adottare strumenti per gestirla. In questo contesto l'intelligenza artificiale si dimostra un alleato imprescindibile per garantire il monitoraggio della rete e la capacità di prendere decisioni adeguate ed efficaci in tempi rapidi. La necessità di sviluppare algoritmi sofisticati e modelli decisionali basati sugli eventi, consentirà infatti a GARR di cogliere le opportunità e affrontare le sfide nel momento in cui si presentano

Keywords. Rete, intelligenza artificiale, digitale, data centre, sdn, fibra ottica

Il bisogno di una nuova rete

Negli ultimi anni è parso quanto più evidente che il modo di fruire della rete è cambiato. Come sappiamo, il web e digitale sono oramai pervasivi nelle nostre vite e spesso parlare di distinzione tra virtuale e reale perde di significato, nelle attività di ricerca così come nella vita di tutti i giorni.

A fronte di questa evoluzione in senso pervasivo del digitale, però, non c'è stato un equivalente sviluppo del concetto di rete, la quale ha risposto alla crescita esponenziale di dati e del loro uso con capacità sempre maggiori senza però modificare nella sostanza il suo modello di funzionamento.

L'attività di aggiornamento della rete GARR ha permesso un ripensamento del suo modello, partendo dalle esigenze crescenti ed eterogenee della sua comunità di utenti. Insieme ad una sempre maggiore capacità, infatti, le necessità della ricerca richiedono una maggiore capillarità sul suolo nazionale, per portare in tutte le aree del paese lo stesso livello di accesso, avendo caratteristiche di flessibilità che permettano l'adattamento nel tempo dell'infrastruttura e l'accesso su richiesta alle risorse di rete e servizi.

GARR-T, una rete fatta come un data centre con un'anima (ottica) aperta

GARR ha voluto dare una soluzione a queste esigenze con la nuova rete GARR-T, in produzione dal 2023 e oggetto di ulteriore potenziamento grazie ai progetti PNRR ICSC e TeRABIT. Si tratta di una rete molto più estesa (20.000 km di fibra ottica, rendendo GARR una delle reti della ricerca più estese in Europa) e con una maggiore capacità, che passa da 3 Tbps a 20 Tbps con l'obiettivo di arrivare fino a 50 Tbps. Eppure la vera novità di questa rete non risiede nella sua capacità, bensì nel modo in cui è stata concepita.

In GARR-T l'architettura è basata sul modello utilizzato per i data centre: ovvero con nodi definiti Spine e Leaf, per l'aggregazione e l'accesso, e nodi Edge, per la terminazione e il

servizio verso l'esterno. Questa scelta progettuale ha permesso di ottenere il meglio delle potenzialità dei mondi delle reti data centre e delle reti di dorsale. Infatti, pur mantenendo uniformità e uguaglianza nei servizi, si favorisce la capillarità e l'estendibilità della rete, garantendo l'agilità nei nodi più piccoli e lasciando l'impatto maggiore (energetico e di spazi) solo sui nodi di core. Per quanto riguarda il modello adottato per la rete ottica, GARR-T si basa su un modello parzialmente disaggregato, nel quale la fibra e gli apparati sono visti come mattoncini differenti che, combinati tra loro, producono l'infrastruttura hardware ottica di trasporto. Questo modello garantisce efficienza e sostenibilità nel tempo perché svincola il ciclo di vita della fibra da quella degli apparati permettendo anche di operare su una rete indipendentemente dal fornitore degli apparati utilizzati, evitando lock-in tecnologici. La piattaforma si basa sull'uso di un sistema di linea aperto (OLS, Open Line System) che insieme alla fibra ottica e ad un sistema software di gestione e controllo, fornisce l'infrastruttura attiva capace di instradare qualunque segnale ottico permettendo la gestione flessibile dello spettro ottico.

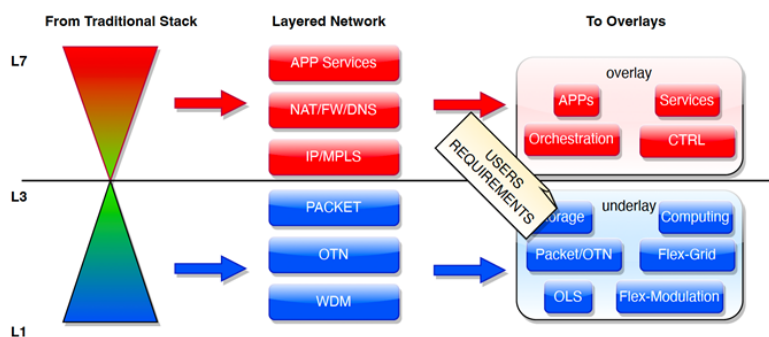
Dall'automazione ad un nuovo paradigma

Gestire un'infrastruttura di rete su scala geografica con queste caratteristiche significa gestire un ambiente maggiormente complesso che in passato. Per questo è stato previsto un alto livello di automazione, per assicurarne affidabilità e misurabilità attraverso dei sistemi di monitoraggio avanzati. I processi, che nella rete precedente (GARR-X) erano condotti in modo semi manuale, in GARR-T vengono completamente rivisti e viene adottato un modello di astrazione che consente di vedere la rete secondo le sue funzioni e non più come accesso alle singole componenti, il tutto in tempo reale. Questo modello è molto simile a quello in uso nel mondo ICT e cloud, ovvero il modello per intenti.

Si tratta di un vero e proprio cambio di paradigma nel modo in cui viene intesa la rete, che passa da un modello a clessidra, che divide la rete in ciò che si trova al di sopra o al di sotto del livello IP, a un modello composto da un underlay (hardware) su cui poggia l'overlay (software). Raccogliendo i dati di tutte le singole componenti della rete, partendo dalla parte infrastrutturale fino alla parte applicativa (o IT), è possibile pensare alla rete come un unico ecosistema in cui le parti, tra loro interconnesse, concorrono all'erogazione dei servizi. Si tratta di un cambio di prospettiva che vede la rete come parte integrante dell'erogazione dei servizi e non più come semplice strumento per mettere in comunicazione gli utenti con i servizi.

Ma come poter gestire una tale mole di dati e intervenire quando necessario?

Fig. 1
Passaggio da un modello a clessidra (layered network) a un modello disaggregato (bricks, mattoncini)



L'intelligenza artificiale come strumento per decidere al tempo degli eventi

In questo contesto le tecniche di intelligenza artificiale e machine learning si dimostrano degli alleati fondamentali. Sfruttando le capacità di calcolo dei data centre GARR e attraverso gli strumenti di telemetria su tutte le componenti della rete è possibile sfruttare gli algoritmi per interpretare i dati in tempo reale, individuare modelli, tendenze e correlazioni altrimenti non facilmente visibili, e intervenire prontamente nel caso di necessità. Si tratta di un modello decisionale per eventi (event driven) ed è la chiave per poter intervenire prontamente laddove il comportamento della rete (o di un servizio) si discosta dal comportamento atteso (malfunzionamento), aumentando così i livelli di affidabilità (reliability) del sistema.



Fig. 2
Nuove metriche - Strumenti Automatici in grado di realizzare e gestire il servizio

L'obiettivo di lungo periodo è raggiungere le cosiddette self-driving network (SDN), ossia le reti a "guida autonoma" in grado non solo di segnalare malfunzionamenti e analizzare le possibili cause, ma anche di valutare autonomamente quali adattamenti suggerire agli operatori, o apportare direttamente, allo scopo di garantire la qualità richiesta dai servizi utente.

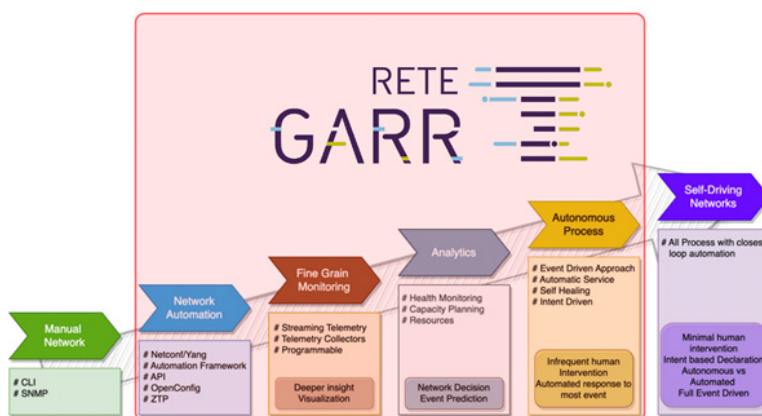


Fig. 3
il percorso di automazione di GARR-T verso le self-driving network

K.Kompella, Self-Driving Networks (2017)

La virtualizzazione dei servizi

In questo nuovo ecosistema i servizi sono più affidabili e viene anche semplificata la loro creazione grazie alla tecnologia di virtualizzazione a container, che scompone il processo in tanti agili micro-servizi. GARR-T può contare su una infrastruttura cloud distribuita e nei vari nodi della rete vengono creati dei mini data centre nei quali ci sono e ci saranno sempre di più sia i servizi infrastrutturali fondamentali (monitoring, sicurezza, AAI, strumenti di analitica) sia i servizi applicativi (eduroam, GARRBox, Mirror,...).

Non solo networking – un mix di competenze

Per poter creare un sistema che controlli la nuova rete GARR dal livello infrastrutturale fino a quello applicativo è fondamentale riuscire a parlare le diverse lingue del sistema, dagli apparati fino al codice degli algoritmi. Per questo il fattore umano è determinante per raggiungere il mix di competenze necessario a costruire un sistema fondato su soluzioni open source e multi vendor gestite interamente da GARR. Tutto questo per configurare la rete su misura delle esigenze della comunità, in continua evoluzione. Il mix di competenze, che vanno dal networking alla componente IT alla capacità di sviluppo, ha permesso di trasformare la rete GARR in un'infrastruttura attiva che riesce a fornire più servizi con un alto livello di affidabilità ad un costo sensibilmente inferiore rispetto al modello di rete precedente.

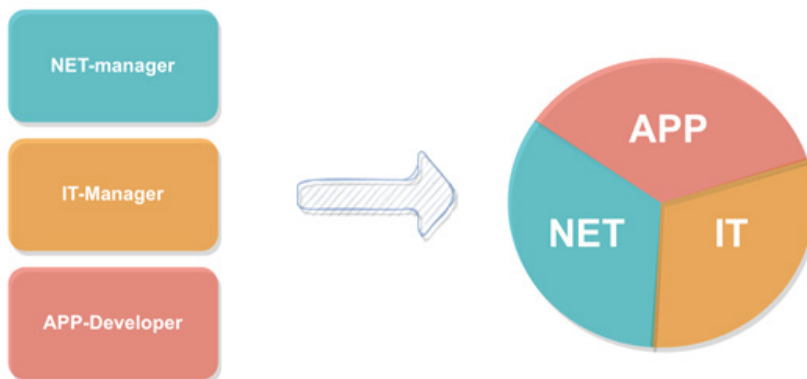


Fig. 4
Servono competenze in tutti i settori del digitale, dal networking, all'IT, alla capacità di sviluppo

Conclusioni

Al di là del miglioramento nell'affidabilità e nel numero dei servizi che GARR offre alla comunità di utenti, operare un'infrastruttura di rete secondo il modello per intenti richiede l'adozione di criteri di disegno, metodologie, così come lo sviluppo di strumenti tecnologici che potranno essere messi a disposizione della comunità GARR.



Autore

Massimo Carboni massimo.carboni@garr.it

Massimo Carboni è vicedirettore e Chief Technical Officer del GARR. Si è laureato in Fisica all'Università degli Studi di Roma La Sapienza e da oltre 20 anni si occupa di calcolo e reti. Nell'ambito del calcolo scientifico, dagli anni 90 si è occupato della transizione dai sistemi di calcolo proprietari a quelli aperti (Unix). Durante questo periodo ha partecipato allo sviluppo di simulazioni software di tipo Montecarlo per la fisica nucleare e subnucleare (HEMAS, FLUKA). Dalla fine degli anni 90 svolge la propria attività nell'ambito del networking acquisendo una notevole esperienza nel campo delle reti ottiche, delle reti a pacchetto e su tematiche infrastrutturali collegate alle reti trasmissive. È stato responsabile della progettazione della rete GARR-G (2002) e successivamente di GARR-X (2009). È stato il coordinatore tecnico del progetto GARR-X Progress (2013-16). Nell'ambito dell'evoluzione di rete europea GEANT ha fatto parte del team di esperti che ha disegnato, progettato l'attuale rete paneuropea Géant. Oggi coordina il progetto GARR-T (Terabit), la nuova infrastruttura di rete nazionale per l'università e la ricerca.

I progetti ICSC, il Centro Nazionale HPC, Big Data e Quantum Computing e TeRABIT

Claudio Grandi¹, Mauro Campanella²

¹INFN-Bologna, ²GARR

Abstract. ICSC è uno dei centri nazionali previsti dal PNRR. Lo scopo di ICSC, iniziato nel settembre 2022, è di creare l'infrastruttura digitale per la ricerca e l'innovazione, sfruttando le infrastrutture HPC, HTC e Big Data esistenti ed evolvendole verso un modello cloud data-lake. L'infrastruttura IT di ICSC è costruita sulle infrastrutture di calcolo esistenti fornite dagli attori principali: GARR, la NREN italiana, fornisce l'infrastruttura di rete, la cui capacità sarà aumentata a multipli Tbps; CINECA che ospita Leonardo, uno dei più potenti sistemi HPC al mondo che sarà ulteriormente potenziato e affiancato da un computer quantistico; l'INFN contribuisce con la sua infrastruttura Big Data distribuita, costruita per rispondere alle esigenze della fisica delle alte energie. TeRABIT, è un progetto sulle Infrastrutture di Ricerca finanziato anch'esso dal PNRR, iniziato nel gennaio 2023, il cui scopo è il potenziamento delle infrastrutture di GARR, CINECA e INFN in totale sinergia con ICSC

Keywords. HPC, HTC, Network, Cloud, Federation

Introduzione

Nei prossimi anni sarà prodotta dal mondo scientifico, industriale e istituzionale una quantità di dati senza precedenti. Il supercalcolo, le simulazioni numeriche, l'intelligenza artificiale, l'analisi dati ad alte prestazioni e la gestione dei Big Data sono essenziali e strategici per dare una risposta alle sfide della società e per stimolare un processo di crescita sostenibile.

Da anni gli USA, la Cina e il Giappone stanno facendo progressi in questo ambito. Anche l'Europa ha sempre dato supporto al calcolo ad alte prestazioni e negli anni recenti ha precisato la propria strategia. I pilastri principali sono le iniziative European HPC Joint Undertaking (EuroHPC JU), la European Open Science Cloud (EOSC), la European Processor Initiative (EPI) e il Quantum Flagship. L'Italia ha finanziato, assieme all'entità legale EuroHPC JU, la creazione di Leonardo, un sistema di supercalcolo pre-exascale, quarto nella Top 500 mondiale, che fornisce servizi ai ricercatori pubblici e privati in Italia e in Europa. Dopo l'emergenza COVID, la EU ha lanciato il programma NextGenerationEU, con lo scopo di facilitare la ripresa e allo stesso tempo trasformare l'economia e la società europea. Degli 806,9 miliardi di euro stanziati, 191,5 sono stati assegnati all'Italia (PNRR), che ne ha dedicati 30,88 alla ricerca e all'educazione; di questi, 11,44 sono per l'iniziativa Dalla ricerca all'impresa e 1,58 per l'iniziativa Infrastrutture di Ricerca.

In risposta a queste azioni, nel 2022 è stata definita la strategia italiana per HPC e big Data, basata sui seguenti punti:

- costruzione di una infrastruttura cloud di supercalcolo combinata con sistemi general purpose;
- creazione di centri di eccellenza per lo sviluppo di applicazioni;
- creazione di forti collegamenti fra il mondo della ricerca e quello industriale;
- istruzione di giovani scienziati e manager per diventare esperti nel settore;
- messa in atto di strategie per l'innovazione e l'aumento del TRL anche per piccole e medie imprese senza competenze specifiche;
- massimizzazione dell'impatto socioeconomico delle iniziative, con attenzione alle implicazioni etiche;
- creazione di un gruppo di esperti con il compito di analizzare e monitorare le implicazioni etiche della gestione dei big data.

Il Centro Nazionale di Ricerca in HPC, Big Data e Quantum Computing (ICSC) è stato finanziato nell'iniziativa Dalla ricerca all'Impresa con 320 milioni di euro. Il progetto Terabit Network for Research and Academic Big Data in Italy (TeRABIT) è stato finanziato con 41 milioni di euro nell'iniziativa Infrastrutture di Ricerca. Lo scopo di ICSC è creare l'infrastruttura digitale per la ricerca e lo sviluppo, partendo dalle infrastrutture HPC, HTC e big data esistenti ed evolvendole verso un modello cloud data-lake accessibile alle comunità scientifiche e industriali. Lo scopo di TeRABIT è il potenziamento delle infrastrutture ricerca digitali GARR-T (GARR), PRACE-Italy (OGS, CINECA) e HPC-BD-AI (INFN) in totale sinergia con ICSC.

1. Organizzazione del progetto ICSC

ICSC è iniziato nel settembre 2022 e la fase di realizzazione e durerà 36 mesi. È organizzato secondo il modello hub & spoke con l'hub responsabile della gestione e validazione del programma di ricerca e gli spoke, con gli affiliati, responsabili per la definizione e l'implementazione del programma.

ICSC include uno spoke trasversale, Spoke-0 Supercomputing Cloud Infrastructure e 10 spoke tematici, come mostrato in figura 1.

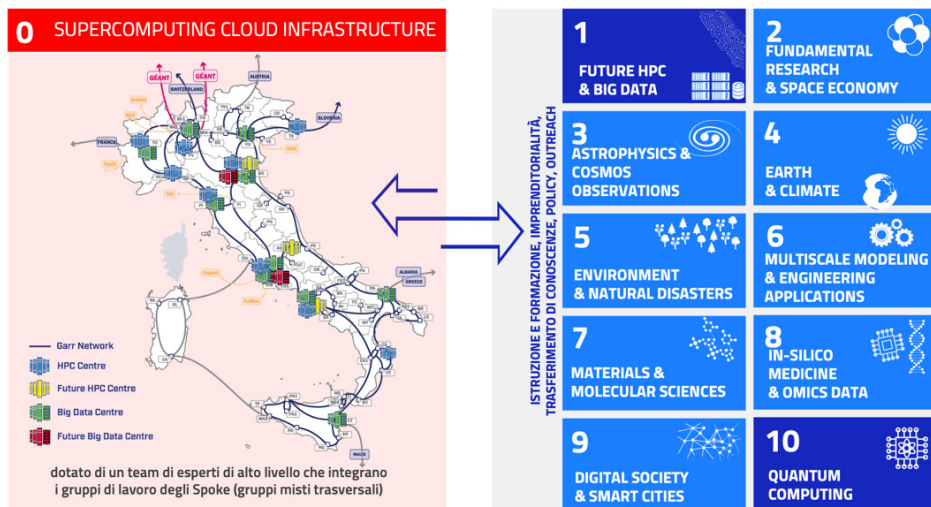


Fig. 1
Gli 11 spoke di ICSC

Lo Spoke-0 si basa sulle infrastrutture dei principali attori nazionali: la rete è fornita da GARR, la NREN italiana, i sistemi HPC sono forniti da CINECA e l'infrastruttura HTC/big Data è fornita dall'INFN. I 10 spoke tematici forniscono la guida e il supporto alle comunità scientifiche e industriali nei diversi campi applicativi. Gli spoke da 2 a 9 sviluppano applicazioni avanzate nei diversi campi che utilizzano l'infrastruttura fornita da Spoke-0, gli spoke 1 e 10 sviluppano nuove tecnologie nell'ambito HPC, Big Data e Quantum Computing.

I partner fondatori della fondazione ICSC sono 25 università, 12 istituti di ricerca e 14 aziende private strategiche; i numeri sono stati limitati dalle prescrizioni del bando e non dall'interesse per l'iniziativa. In totale sono coinvolti circa 1500 ricercatori condivisi dai partner, più di 250 nuovi ricercatori assunti e oltre 250 dottorandi. Parte dei 320 milioni di euro saranno usati per bandi a cascata (32 milioni) e per innovation grants (32 milioni). La sede di ICSC sarà al Tecnopolo di Bologna, dove sono già attivi il centro di calcolo di ECMWF, il nuovo sistema pre-exascale Leonardo di CINECA e dove è in fase di trasferimento il centro di calcolo Tier-1 di INFN-CNAF.

2. Il progetto TeRABIT

Il progetto TeRABIT è iniziato nel gennaio 2023 e durerà 30 mesi. Il progetto ha come scopo il potenziamento di tre infrastrutture di ricerca digitali del piano nazionale: la rete GARR-X (ora GARR-T) di GARR, il sistema HPC Prace-Italy di CINECA e OGS e l'infrastruttura di calcolo distribuita HPC-BD-AI dell'INFN. Non è un caso che le tre infrastrutture siano quasi identiche a quelle dello Spoke-0 di ICSC. In effetti i due progetti sono costruiti per essere totalmente sinergici, anche se con qualche specificità. Gli interventi sulla rete GARR interesseranno regioni differenti; in particolare TeRABIT prevede il potenziamento su fibra ottica dedicata del collegamento con la Sardegna, dove si trova anche uno dei siti candidati ad ospitare l'Einstein Telescope, il futuro osservatorio europeo per le onde gravitazionali. OGS con CINECA potenzieranno il sistema Tier-1 Galileo 100. L'INFN costruirà sistemi HPC eterogenei, le cosiddette HPC bubbles, indirizzati a casi d'uso più vicini ai siti dell'utente e diversi da quelli considerati in ICSC.

3. L'infrastruttura Cloud Supercomputing

3.1 Rete

L'infrastruttura GARR-T evolve la rete nazionale attuale ed è in fase di installazione da parte di GARR. La rete è basata su fibra ottica, attualmente oltre 16000 km, di proprietà di GARR tramite Indefeasible Right of Use (IRU), che utilizza un sistema di trasmissione ad alte prestazioni basato su tecnologia Open Line System, con l'orchestrazione basata su software e hardware sotto il controllo di GARR. GARR-T è in grado di arrivare a trasportare diverse decine di Tbps garantendo un elevato livello di flessibilità, capace di fornire accesso al livello ottico anche a utenti che non si collegano direttamente sui PoP GARR. GARR-T è connessa alla rete europea GÉANT.

3.2 HPC

Il nuovo sistema HPC Tier-1 Leonardo di CINECA, co-finanziato dalla EuroHPC, ad oggi è in grado di fornire circa 250 Pflops. Con l'ulteriore co-finanziamento previsto in ICSC la potenza sarà aumentata del 30%. Inoltre, sarà integrato un sistema di calcolo quantistico, consentendo quindi di combinare le tecnologie digitali e quantistiche per casi d'uso avanzati, in linea con le strategie di EuroHPC. Sarà potenziato il sistema Tier-1 Galileo 100 di PRACE-Italy e costruito un nuovo Tier-1 a Napoli.

3.3 Big Data Federated Cloud

L'INFN sta gestendo un Sistema di 10 centri di calcolo distribuiti sul territorio nazionale, dedicati alla gestione e all'analisi dei dati di oltre 50 esperimenti internazionali. L'attuale capacità sarà raddoppiata entro il 2025, in modo da poter ampliare la base di utenti. Il centro di calcolo principale è in fase di migrazione al Tecnopolo di Bologna. Le infrastrutture degli altri centri saranno potenziate e rese energeticamente più efficienti. Saranno inoltre creati due nuovi centri di calcolo dedicati alla resilienza a disastri naturali ed antropici ai LNGS e alla Space Economy ai LNF. Per poter indirizzare casi d'uso nuovi, si installeranno sistemi eterogenei basati non solo su CPU ma anche su GPU, FPGA e altre architetture. L'architettura globale è basata sulla federazione a diversi livelli (IaaS, PaaS, SaaS) delle risorse dei centri attuali. Il modello è dinamico e consente l'inclusione di risorse e utenti in base alle esigenze e alle disponibilità. I sistemi saranno compatibili con gli standard definiti da EuroHPC ed EOSC.

4. Conclusioni

Il finanziamento sul calcolo del PNRR è una opportunità unica per potenziare l'infrastruttura di calcolo italiana della ricerca e per intensificare la collaborazione con il sistema produttivo nazionale.

La forte collaborazione fra gli attori principali (rete, HPC e HTC/Big Data) aumenterà l'efficienza e l'efficacia dell'infrastruttura ma sarà necessario aumentare la flessibilità e l'automazione nella gestione dei sistemi.

L'iniezione di un grande numero di nuovi ricercatori e tecnologi che costruiranno la propria carriera sull'analisi dei dati attraverso il calcolo ad alte prestazioni è importante per coprire aree al momento scoperte, anche se l'integrazione nelle attività ordinarie degli enti costituisce una sfida.

Ringraziamenti

Le attività sono finanziate dai progetti PNRR ICSC (M4, C2, I1.4, CP CN00000013, CUP I53C21000340006) e TeRABIT (M4, C2, I3.1, CP IR00000022, CUP I53C21000370006).

Autori

Claudio Grandi Claudio.Grandi@bo.infn.it

Claudio Grandi è dirigente tecnologo all'INFN. Ha iniziato l'attività nel 1989 come ricercatore in fisica delle alte energie. Dai primi anni 2000 il fuoco si è spostato sul calcolo. Ha avuto ruoli di coordinamento nell'esperimento CMS del CERN, nei progetti europei EGEE e nel calcolo dell'INFN. È stato

Presidente della Commissione Calcolo e Reti dell'INFN. È membro del CTS di GARR. Ha contribuito alla preparazione dei progetti PNRR ICSC, TeRABIT e ECOSISTER ed è co-leader di Spoke-0 di ICSC.

Mauro Campanella Mauro.Campanella@garr.it

Mauro Campanella è primo tecnologo all'INFN. Laureato in fisica nel 1985 si è da subito dedicato al calcolo e reti. È uno dei proponenti della prima versione della rete GARR ed ha partecipato e coordinato vari progetti europei di innovazione nelle reti, in particolare la serie GÉANT della dorsale europea. Attualmente è responsabile scientifico del progetto TeRABIT.