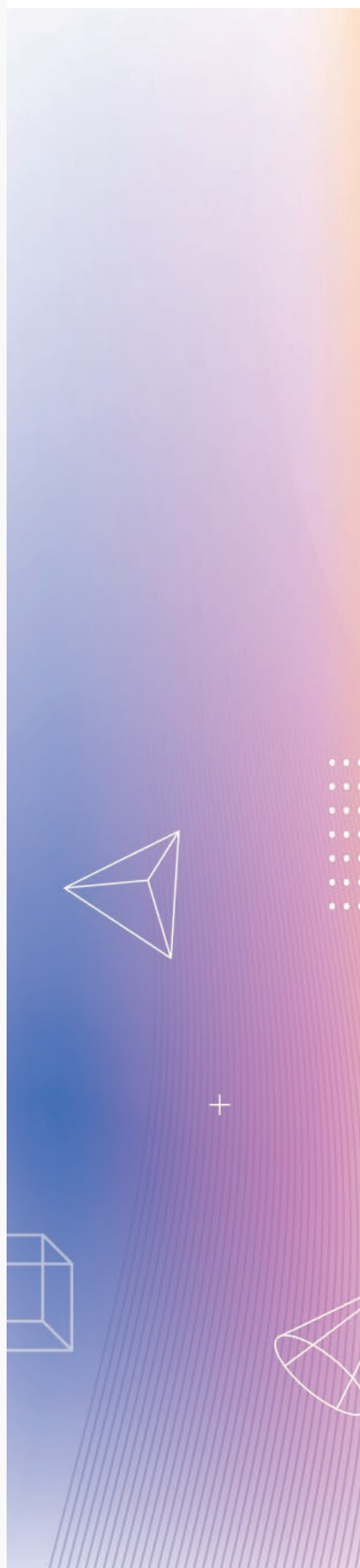


CONFERENCE  
GARR  
2024

# NAVIGARE LA COMPLESSITÀ

Infrastrutture e  
competenze digitali  
per la ricerca

SELECTED  
PAPERS



CONFERENCE  
GARR  
2024

# Navigare la complessità

Infrastrutture e competenze digitali  
per la ricerca

Brescia, 29-31 maggio 2024

SELECTED  
PAPERS





Conferenza GARR 2024 - Navigare la complessità - Infrastrutture e competenze  
digitali per la ricerca  
Brescia, 29-31 maggio 2024

ISBN 978-88-946629-3-1

DOI 10.26314/GARR-Conf24-proceedings

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



Editore: Associazione Consortium GARR

Via dei Tizii, 6, 00185 Roma, Italia

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

Curatori editoriali: Marta Mieli, Carlo Volpe

Progetto grafico: Carlo Volpe

Impaginazione: Marta Mieli

Prima stampa: Ottobre 2024

Numero di copie: 400

Stampa: Tipografia Graffietti Stampati snc

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

Tutti i materiali relativi alla Conferenza GARR 2024 sono disponibili all'indirizzo:  
[conf24.garr.it/](http://conf24.garr.it/)

# Indice

- 8 Dal mito alla realtà, visione strategica e opportunità tecnologica  
Massimo Carboni
- 13 Navigare la Complessità dell'Intelligenza Artificiale attraverso la Formazione e la  
Condivisione Multidisciplinare: l'esperienza di OpenAIF  
Vivaldo Moscatelli, Emanuela Torcinaro, Massimo Conte
- 18 Natural Language Processing e Computer Vision per l'accessibilità dei video di lezioni con  
formule  
Tiziana Armano, Gabriele Baratto Cristian Bernareggi, Sara Bertoldo, Erika Brunetto,  
Anna Capietto, Manuela Caramagna, Sandro Coriasco, Nazareno Defrancesco, Mattia  
Ducci, Maria Luisa Gabrielli, Mariafrancesca Guadalupi, Alessandro Mazzei, Angelo Saccà,  
Francesco Tarasconi
- 24 Artificial Intelligence, LLM and RAG supporting Library and Information Science:  
challenges and opportunities  
Giorgia Di Marcantonio
- 29 Il patrimonio culturale come sistema di relazioni. Verso un nuovo Sistema del Catalogo  
Nazionale  
Carlo Birrozzi, Fabrizio Magnani, Elena Musumeci, Chiara Veninata
- 36 Geoportale Nazionale per l'Archeologia: la carta archeologica del territorio nazionale alla  
prova della sostenibilità  
Valeria Boi, Ada Gabucci, Francesco Marucci
- 41 La Piattaforma di big data in Istat nei processi statistici e di sperimentazione  
Francesco Altarocca, Marco Polizzi
- 47 On Data Throughput in Real-Time Structural Health Monitoring: a Next-Generation EU  
funded Project use case  
Paolo Bellagente
- 52 L'utilizzo delle infrastrutture digitali per la riqualificazione del patrimonio edilizio  
italiano: un caso studio  
Giacomo Bizzarri, Enrica Boldrin, Laura Ferrari

- 59 **Data, Reduced Order Models and Artificial Intelligence for Digital Twin**  
Gianluigi Rozza, Laura Meneghetti, Andrea Martini
- 66 **Una infrastruttura Cloud federata per ICSC e TeRABIT**  
Claudio Grandi
- 71 **Progetti di comunicazione Quantistica nella Regione Friuli Venezia Giulia**  
Angelo Bassi, Giorgio Giorgetti
- 76 **GPU Offloading in MLOps: Navigating the Multicloud Ecosystem for Flexible AI/ML Deployments**  
Fulvio Rizzo, Giuseppe Zangari
- 81 **Strategie per la progettazione della ricerca finalizzate alla fruizione e valorizzazione dei beni culturali**  
Sara Alimenti, Elena Gentilini
- 86 **Un cruscotto di monitoraggio della ricerca basato su dati aperti per l'Università degli Studi di Milano**  
Dario Basset, Stefano Bolelli Gallevi
- 90 **Itinerari Digitali**  
Cecilia Perrone
- 96 **Sovranità digitale: DPO, protezione dei dati e cybersicurezza**  
Lucio Badiali
- 100 **NeuroArtP3: un modello di gestione del trattamento dei dati nell'ambito di un Programma di rete dopo l'entrata in vigore del GDPR**  
Manuela Marengo, Irene Sanginisi, Matteo Pellini, Andrea Gambino, Luisa Franco, Cristina Grandi, Marco Damonte Prioli, Luca Boni, Antonio Uccelli
- 106 **Il portale delle Raccolte Fotografiche: un progetto di condivisione dei patrimoni e delle conoscenze tra pubblico e privato**  
Alessandro Coco



**CONFERENZA GARR 2024**  
Brescia, 29-31 maggio

# Navigare la complessità

Infrastrutture e competenze digitali per la ricerca

## Comitato di programma

Riccardo Adamini - Università di Brescia  
Claudio Allocchio - GARR  
Davide Bacciu - Università di Pisa  
Claudia Battista - GARR  
Elis Bertazon - GARR  
Alberto Bruni - Ministero della Cultura  
Mauro Campanella - GARR  
Massimo Carboni - GARR  
Luigi Cattivelli - CREA  
Ilaria Comelli - Università di Parma  
Francesca Di Donato - CNR-ILC  
Sara Di Giorgio - GARR  
Fabio Farina - GARR  
Carlos Fernandez - Human Technopole  
Fulvio Galeazzi - GARR  
Leonardo Lanzi - GARR  
Alberto Masoni - INFN  
Marta Mieli - GARR  
Gabriella Paolini - GARR  
Maria Chiara Pievatolo - Università di Pisa  
Simon Pietro Romano - Università di Napoli Federico II  
Stefano Salon - OGS  
Sabrina Tomassini - GARR  
Davide Vaghetti - GARR  
Chiara Veninata - Ministero della Cultura  
Carlo Volpe - GARR

Tutte le presentazioni e maggiori informazioni  
sono disponibili sul sito dell'evento:  
[conf24.garr.it/it/](https://conf24.garr.it/it/)

# Dal mito alla realtà, visione strategica e opportunità tecnologica

Massimo Carboni

GARR

**Abstract.** La tecnologia digitale, potenziata dall'AI, è dirompente. Di fronte a tanta forza ed una capacità di calcolo crescente, viene spontaneo affidarsi ad essa come ad un oracolo per trovare soluzioni alla complessità delle sfide di oggi. Nonostante un contesto internazionale di forte asimmetria e polarizzazione delle risorse, è importante che la comunità accademica e della ricerca continui a creare conoscenza, sviluppando nuove competenze e mantenendo alta la consapevolezza sugli strumenti utilizzati.

**Keywords.** Tecnologia digitale, AI, digitalizzazione, data centre, deep learning

## 1. Introduzione

L'avvento dell'Intelligenza Artificiale e, in particolare dell'AI Generativa, ha avuto un impatto travolgente nell'era digitale, promettendo immense opportunità ma sollevando anche cruciali interrogativi etici e pratici. Questa trasformazione epocale richiede una profonda riflessione su come bilanciare i potenziali benefici dell'AI con i rischi di una delega eccessiva della capacità decisionale umana a pochi soggetti tecnologici dominanti. Questo paper mira a delineare un approccio per mantenere il ruolo guida dell'essere umano in questa nuova era tecnologica, assicurando che l'AI rimanga uno strumento al servizio dell'umanità e non un fine in sé.

## La conoscenza come motore dell'evoluzione umana

Il progresso dell'umanità è stato guidato dallo sviluppo cognitivo, dalla curiosità e dalla creatività che ci contraddistinguono come specie. Questi tratti ci hanno spinto a sviluppare strumenti e tecnologie per soddisfare i nostri bisogni e migliorare le nostre condizioni di vita. Tuttavia, il vero motore propulsivo dell'evoluzione umana risiede nell'apprendimento sociale, nella comunicazione e nella collaborazione. È stata la capacità di condividere la conoscenza, di tramandarla e di costruire collettivamente che ha permesso all'umanità di compiere progressi senza precedenti. Un esempio emblematico è la scoperta del fuoco, un'innovazione che ha rivoluzionato la vita dell'uomo primitivo e che, nel corso dei millenni, è stata compresa, trasformata e perfezionata fino a condurre alle più moderne tecnologie energetiche.

La tecnologia digitale e in particolare l'intelligenza artificiale rappresenta il tentativo più

recente e finora dirompente di utilizzare tutta la conoscenza collettiva dell'uomo per dare una risposta a problemi che sono sempre più collettivi e di respiro globale, dai problemi di sicurezza (cyber e non), alle catene di approvvigionamento delle risorse a minacce per la stessa esistenza della specie, come il cambiamento climatico o l'antibiotico resistenza. Affinché si possa continuare ad evolvere ed affrontare i grandi temi che riguardano l'umanità è necessario che questo motore alla base dell'evoluzione umana rimanga ben vivo anche e soprattutto nell'era dell'AI, in modo che la tecnologia continui ad essere una risposta ai bisogni dell'uomo e non diventi uno strumento per creare nuovi bisogni.



### **La tempesta perfetta: complessità e semplificazione**

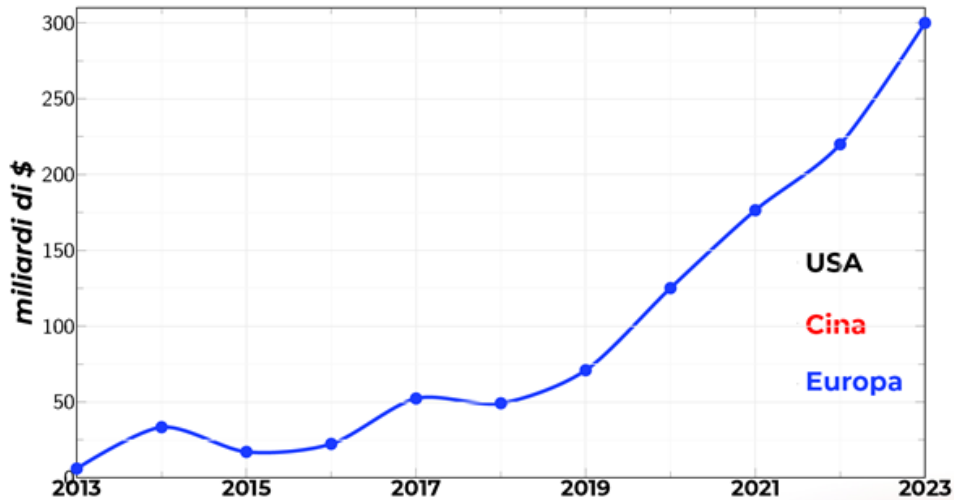
Nell'attuale panorama digitale, stiamo affrontando quella che potremmo definire una "tempesta perfetta". Da un lato il mondo si fa sempre più complesso: per affrontare problemi su scala globale il processo di digitalizzazione sta procedendo in modo disordinato e non strutturato, con una moltitudine di attori che perseguono simultaneamente obiettivi individuali, aumentando esponenzialmente la complessità del sistema. Dall'altro lato, vi è la tendenza a semplificare il sistema attraverso la tecnologia, confidando in particolare nell'AI generativa come mezzo per semplificare questa complessità, alla stregua di un oracolo. Ma proprio come un oracolo, i meccanismi e le basi dati su cui poggia questa tecnologia non sono trasparenti o accessibili agli utilizzatori, che così si trovano a delegare l'onere della decisione a soggetti terzi. Di fronte a questo scenario appare evidente che l'utopia iniziale che portò alla nascita di Internet come sistema neutrale, aperto e paritario che rimuove gli intermediari dall'accesso al sapere, si sta rapidamente trasformando in una realtà in cui pochi giganti tecnologici detengono un potere sempre maggiore, di fatto diventando nuovi intermediari del sapere.

Questa situazione rischia di scivolare verso un contesto distopico in cui la capacità decisionale viene delegata a una manciata di soggetti che controllano l'AI. Rimanendo in

ambito di metafora, ciò che si prefigura è un futuro in cui il golem, l'automa privo di vera intelligenza ma dotato di una forza sovrumana creato per aiutare l'essere umano, diventa autonomo e sfugge al controllo del suo creatore.

## Investimenti globali e potenziali vantaggi dell'AI

Gli investimenti globali nell'AI hanno raggiunto cifre record, con gli Stati Uniti e la Cina in testa e l'Europa che insegue a distanza. Questi investimenti massicci stanno alimentando un'accelerazione senza precedenti nello sviluppo dell'AI.



Fonte: Our World in Data

I potenziali vantaggi dell'AI sono molti e promettenti. Si può prevedere un aumento significativo dell'efficienza e della produttività in vari settori, accompagnato da un miglioramento dei processi decisionali grazie all'analisi di grandi quantità di dati. L'AI promette inoltre di offrire una personalizzazione avanzata dei servizi, aprendo la strada a nuove scoperte e innovazioni in campi come la medicina e la ricerca scientifica. L'ottimizzazione delle risorse potrebbe rappresentare un aiuto per la sostenibilità ambientale, mentre miglioramenti sostanziali sono attesi nei settori della sanità, dell'educazione e dei trasporti. Alcuni esempi di questi vantaggi li abbiamo già vissuti. Durante la pandemia di COVID-19 l'AI ha svolto un ruolo cruciale nello studio di potenziali molecole terapeutiche, accelerando il processo di scoperta grazie alla sua capacità di simulare e scartare scenari non promettenti.

## Rischi ed effetti collaterali

Nonostante i potenziali benefici, l'adozione diffusa dell'AI comporta anche rischi significativi e effetti collaterali che non possono essere ignorati. Innanzitutto è importante sottolineare che l'AI generativa sia una tecnologia estremamente energivora a causa del gran numero di data centre necessari al suo funzionamento. Da un punto di vista ambientale,

la sostenibilità dei grossi sistemi di deep learning (LLM) è in discussione al punto che, a causa dell'aumento delle emissioni di CO2 prodotta, grossi player come Google hanno dichiarato che difficilmente potranno raggiungere gli obiettivi climatici di carbon neutrality del 2030. Inoltre, uno dei rischi più pressanti è sicuramente la polarizzazione delle risorse e del potere decisionale nelle mani di pochi attori tecnologici dominanti. Questa concentrazione di potere è accentuata dalle enormi barriere all'ingresso rappresentate dagli investimenti necessari per sviluppare e addestrare modelli AI avanzati.

Sul fronte normativo, le leggi e le regolamentazioni faticano a tenere il passo con la rapida evoluzione tecnologica, creando vuoti legislativi potenzialmente pericolosi.

Le implicazioni geopolitiche dell'AI sono altrettanto rilevanti, con le infrastrutture digitali che stanno ridisegnando gli equilibri di potere globali.

E poi c'è un altro effetto collaterale particolarmente preoccupante: il rischio dell'omologazione intesa come perdita della capacità di creare nuovo sapere e di diventare eccessivamente dipendenti dall'AI, che viene intesa come fonte di conoscenza. Una fonte, per di più, intrisa di bias cognitivi provenienti dai materiali forniti all'AI generativa per l'addestramento e sui quali non vi è trasparenza. Di conseguenza, se utilizzata senza consapevolezza sulla fase di addestramento, l'AI rischia di perpetuare e amplificare stereotipi e discriminazioni con notevoli ripercussioni in ambito sociale.

## **Mantenere il controllo**

L'AI generativa può portare tanti benefici ma i rischi vanno presi in considerazione. Per scongiurare il rischio di omologazione e continuare ad essere rilevanti nella creazione di nuovo sapere e di innovazione, è necessario che la comunità accademica e della ricerca, e GARR con essa, partecipino allo sviluppo di questa tecnologia, investendo tempo e risorse per poter utilizzare l'AI in modo consapevole ma anche economicamente sostenibile.

È necessario partecipare allo sviluppo tecnologico nonostante l'attuale asimmetria di risorse in gioco con i big player. Per fare ciò è necessario dotarsi degli strumenti che permettendo di mantenere l'autonomia, ossia la capacità di porsi nuove domande per immaginare nuove risposte e nuove soluzioni, oltre alle richieste del mercato e con la lungimiranza che solo chi pensa al futuro può avere.

Un modo per mantenere il controllo della tecnologia, sfruttando al contempo le capacità dell'AI generativa può essere quello di sviluppare SLM (Small Language Model) con dati interni per sviluppare la knowledge base, una soluzione che è più facilmente controllabile rispetto ad un LLM, sebbene richieda dati di maggiore qualità per l'addestramento e un lavoro maggiore da parte dei data scientist. Nel caso di GARR, una soluzione di questo tipo permetterebbe di sfruttare i molti dati a disposizione sulla rete per ottimizzarla, coltando e mantenendo le competenze necessarie per il suo funzionamento.

A livello di comunità della ricerca, però, è necessario il coinvolgimento di tutti gli attori in gioco, dallo stato alle università al sistema paese e alle collaborazioni con le industrie. A livello europeo è importante collaborare per avere strumenti che aiutino la ricerca europea e non la frenino.

## Comprendere per innovare

Affinché la comunità a della ricerca continui a contribuire alla conoscenza e a fare innovazione è fondamentale cogliere l'opportunità di far parte di questa tecnologia da protagonisti, fungendo da argine all'oligopolio delle grandi aziende al suo sviluppo.

L'idea è quella di continuare a creare innovazione e tecnologia, investendo sulle competenze e sviluppando soluzioni che riescano a sfruttare il valore aggiunto dell'AI senza delegare l'onere della decisione. Ritornando alla metafora del golem: è importante mantenere la conoscenza (emet) per poter controllare la tecnologia.

### Autore



**Massimo Carboni** [massimo.carboni@garr.it](mailto: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 30 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.

# Navigare la Complessità dell'Intelligenza Artificiale attraverso la Formazione e la Condivisione Multidisciplinare: l'esperienza di OpenAIF

Vivaldo Moscatelli<sup>1</sup>, Emanuela Torcinaro<sup>2</sup>, Massimo Conte<sup>3</sup>

<sup>1</sup>Ecosistema Formazione Italia (EFI), <sup>2</sup>Associazione Italiana Formatori (AIF),

<sup>3</sup>Complexity Education Project (CEP)

**Abstract.** Questo studio esplora l'efficacia di #OpenAIF, una piattaforma innovativa per l'apprendimento e la condivisione di conoscenze sull'intelligenza artificiale. Attraverso un approccio metodologico misto, la ricerca analizza come #OpenAIF navighi la complessità dell'IA, promuovendo la multidisciplinarietà e la consapevolezza digitale. I risultati preliminari evidenziano l'impatto significativo della piattaforma nell'incrementare la comprensione dell'IA, stimolare l'interdisciplinarietà e creare una comunità digitale attiva. #OpenAIF emerge come un sistema complesso adattivo, capace di rispondere efficacemente alle sfide del settore. Lo studio sottolinea l'importanza di tali iniziative nell'era digitale e suggerisce direzioni per future ricerche, mirando a ottimizzare l'educazione sull'IA e il suo impatto sulla società

**Keywords.** Intelligenza artificiale, Apprendimento collaborativo, Multidisciplinarietà

## Introduzione

L'avvento dell'intelligenza artificiale (IA) ha portato con sé una rivoluzione tecnologica senza precedenti, caratterizzata da una complessità e una velocità di innovazione che pongono sfide significative sia per gli esperti del settore che per il grande pubblico. In questo contesto, il progetto #OpenAIF ([www.bit.ly/openaif](http://www.bit.ly/openaif)) si propone di offrire un approccio innovativo all'orientamento e alla comprensione di queste tecnologie emergenti.

#OpenAIF (#OA) nasce con l'obiettivo di colmare il divario tra la conoscenza tecnologica avanzata e la formazione accessibile, promuovendo un modello di apprendimento collaborativo che attraversa discipline e contesti diversi.

## 1. Focus della ricerca

Il nostro studio si concentra sull'analisi delle strategie implementate per affrontare e decifrare la complessità nell'adozione e nell'integrazione dell'intelligenza artificiale nelle pratiche formative. In particolare, esaminiamo come la piattaforma faciliti l'accesso e la condivisione di conoscenze e risorse tra diversi ambiti disciplinari, spaziando dal settore scientifico alla produzione musicale.

Un aspetto chiave della nostra indagine è l'approccio multidisciplinare promosso da #OA nel trattare temi di frontiera come cybersecurity, big data e machine learning. Questo approccio non solo arricchisce il panorama formativo, ma contribuisce anche a creare ponti

tra discipline apparentemente distanti, favorendo una comprensione più olistica dell'IA e delle sue applicazioni.

Particolare attenzione è rivolta al ruolo della piattaforma nel facilitare il trasferimento tecnologico nel contesto della ricerca e dell'istruzione, delineando così nuove prospettive per la formazione e la divulgazione scientifica nell'era digitale.

## 2. Metodologia

Per approfondire il ruolo e l'efficacia di #OA nel navigare la complessità dell'integrazione dell'IA nelle pratiche formative, abbiamo adottato un approccio metodologico misto (vedi <https://bit.ly/numerioopenaif>).

Questo include:

- **Analisi qualitativa:** Abbiamo condotto un'analisi approfondita di interviste, workshop e sessioni di formazione organizzate in #OA. L'obiettivo era comprendere la percezione e l'esperienza dei partecipanti riguardo l'apprendimento dell'IA.
- **Analisi quantitativa:** Abbiamo effettuato un'analisi dei dati di partecipazione e interazione all'interno della piattaforma digitale di #OA. Questo ci ha permesso di misurare l'effettivo impatto sulla comunità di apprendimento, tracciando metriche come il numero di partecipanti attivi e la frequenza di interazione.

Questo approccio metodologico misto ci ha consentito di tracciare un quadro complessivo delle dinamiche di apprendimento e di condivisione della conoscenza generate da #OA, delineando così le pratiche migliori e le sfide ancora aperte nel campo della formazione sull'IA.

## 3. Risultati Preliminari

I risultati preliminari del nostro studio evidenziano l'impatto significativo che #OA ha avuto sulla comunità di apprendimento e formazione nel campo dell'intelligenza artificiale. Possiamo suddividere questi risultati in diverse categorie chiave:

- **Incremento della consapevolezza e comprensione dell'IA:** Si è osservato un notevole aumento della comprensione dell'intelligenza artificiale tra i partecipanti. Questo risultato è stato ottenuto grazie alla diffusione di materiali didattici di alta qualità, alla realizzazione di incontri con esperti del settore (vedi Il calendario degli eventi <https://bit.ly/calendario-openaif>) e all'organizzazione di laboratori pratici sull'utilizzo degli strumenti di IA. Questo approccio ha favorito una maggiore apertura verso l'adozione di tecnologie IA nelle pratiche quotidiane e professionali, superando barriere iniziali di diffidenza o incomprensione.
- **Promozione dell'interdisciplinarietà:** l'analisi qualitativa ha messo in luce l'efficacia della piattaforma #OA nel promuovere l'interdisciplinarietà. Gli utenti hanno sperimentato e apprezzato l'opportunità di condividere conoscenze e prospettive con professionisti e appassionati di settori diversi. Questo scambio ha portato a un arricchimento significativo del bagaglio culturale e professionale dei partecipanti, favorendo approcci innovativi e creativi all'applicazione dell'IA in vari campi.
- **Creazione di una comunità digitale attiva:** l'approccio collaborativo di #OA ha stimola-

to la creazione di una comunità digitale attiva e impegnata. In questo ecosistema, i partecipanti non solo acquisiscono conoscenze, ma contribuiscono anche attivamente con contenuti originali e discussioni. Questo ha promosso un circolo virtuoso di apprendimento e condivisione, dove la conoscenza fluisce in modo multidirezionale tra tutti i membri della comunità.

- #OA come sistema adattivo complesso: uno degli aspetti più interessanti emersi dallo studio è la natura di #OA come sistema adattivo complesso. La piattaforma può essere descritta come un organismo autopoietico, caratterizzato da processi di creazione di conoscenza (attraverso webinar pubblici e condivisione nella community su Discord), trasformazione (gruppi di lavoro su specifici topic autoselezionati e creati dai partecipanti) e dissoluzione di componenti (intesa come thread a cui poi non viene dato seguito).
- Fenomeni di emergenza: attraverso l'interazione tra i partecipanti, è stato possibile osservare fenomeni di emergenza, ovvero proprietà appartenenti al sistema (livello superiore) ma che non appartengono ai singoli componenti. Esempi di questi fenomeni includono la creazione di paper (sottoposti e, in parte, selezionati da Unesco, Conferenza di Parigi, GARR, Protezione Civile) e proposizione di idee innovative frutto della collaborazione nei gruppi di lavoro e l'intelligenza collettiva risultante dalla continua segnalazione di risorse da parte dei partecipanti della community.
- Risposta alla complessità: in linea con la Legge di Ashby, #OA risponde all'aumento di complessità dell'ambiente esterno (il nuovo mondo basato sull'IA, caratterizzato da varietà, variabilità, interdipendenza e indeterminazione) con una complessità interna della community. Questa si manifesta come capacità di adattarsi all'ambiente esterno, attraverso una selezione di risorse significative sull'AI generativa (articoli, bibliografie, siti, corsi, podcast).
- Sistema aperto e decentralizzato: #OA si configura come un sistema aperto, che scambia costantemente informazioni con l'esterno. Allo stesso tempo, è un sistema decentralizzato, frutto dell'attività della rete dei partecipanti attivamente coinvolti. Questa struttura favorisce la resilienza e l'adattabilità della piattaforma, permettendole di evolversi in risposta alle mutevoli esigenze del campo dell'IA.

#### 4. Discussione

I risultati preliminari emersi dallo studio su #OA mostrano un nuovo paradigma di apprendimento e condivisione della conoscenza nel campo dell'intelligenza artificiale. La piattaforma si distingue per la sua capacità di navigare efficacemente la complessità intrinseca dell'IA, offrendo un modello replicabile per iniziative simili in altri campi tecnologici e non solo, emergenti.

L'approccio multidisciplinare di #OA si è rivelato particolarmente efficace nel superare le barriere tradizionali tra diverse aree di expertise. Questo non solo arricchisce il processo di apprendimento, ma prepara anche i partecipanti ad affrontare le sfide reali del mondo IA, dove la convergenza di diverse discipline è spesso la chiave per soluzioni innovative.

La natura di sistema complesso adattivo di #OA merita un'attenzione particolare. Questa caratteristica permette alla piattaforma di evolversi organicamente in risposta alle esi-

genze della comunità e ai rapidi cambiamenti nel campo dell'IA. La capacità di auto-organizzazione e adattamento è cruciale in un settore in così rapida evoluzione come quello dell'intelligenza artificiale.

I fenomeni di emergenza osservati all'interno della comunità #OA sottolineano il potenziale dell'intelligenza collettiva quando applicata all'apprendimento dell'IA. Questi risultati suggeriscono che piattaforme simili potrebbero essere strumenti potenti non solo per la diffusione di conoscenze esistenti, ma anche per la generazione di nuove idee e approcci innovativi.

## 5. Prospettive Future

I risultati preliminari emersi dallo studio su #OA illuminano il percorso verso una più profonda comprensione e integrazione dell'intelligenza artificiale nel tessuto educativo e professionale. Tuttavia, il cammino per navigare efficacemente la complessità dell'era digitale è ancora lungo e richiede ulteriori sforzi di ricerca e sviluppo.

In particolare, la ricerca futura dovrebbe concentrarsi su:

1. L'identificazione di strategie innovative per l'educazione e la formazione in IA, garantendo che queste siano accessibili, eque e in grado di rispondere efficacemente alle esigenze di una società in rapida evoluzione.
2. L'analisi dell'impatto a lungo termine di piattaforme come #OA sulla promozione di una consapevolezza digitale tra i cittadini e sulla preparazione di professionisti capaci di operare in un contesto sempre più tecnologico.
3. Lo studio di come integrare le lezioni apprese da #OA in contesti educativi formali, come università e programmi di formazione aziendale.
4. L'esplorazione di metodi per scalare l'approccio di #OA, mantenendo al contempo la qualità dell'interazione e dell'apprendimento.

## 6. Conclusione

#OA rappresenta un modello promettente per navigare la complessità dell'intelligenza artificiale attraverso la formazione e la condivisione multidisciplinare. Il suo successo sottolinea l'importanza di approcci innovativi e collaborativi nell'educazione tecnologica. Mentre ci muoviamo verso un futuro sempre più permeato dall'IA, iniziative come #OA saranno cruciali per garantire che la società nel suo complesso possa beneficiare di queste tecnologie emergenti, promuovendone, al contempo, un'adozione etica e consapevole.

### Autori



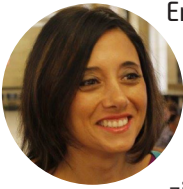
Vivaldo Moscatelli [moscavi@gmail.com](mailto:moscavi@gmail.com)

Informatico per estrazione ma formatore per vocazione, si occupa di didattica digitale dalla fine degli anni 80.

Advisor di Ecosistema Formazione Italia, è alla Scuola di Formazione della Banca d'Italia. Ambassador della Comunità Europea nel progetto European Digital Skills Certificate.

Dai sistemi esperti degli anni 90, nel 2023 ha creato il social #OpenAIF per favorire la diffusione dei temi dell'intelligenza artificiale.

Ha scritto di IA per Franco Angeli, Persone&Conoscenze e AIF Learning News



**Emanuela Torcinaro** [emanuelatorcinaro@gmail.com](mailto:emanuelatorcinaro@gmail.com)

Umanista con dottorato in Letterature e Culture Moderne Comparete tra Italia e Francia, vanta oltre 20 anni di esperienza come insegnante e formatrice. Specializzata in e-learning, progetta percorsi digitali centrati sul discente. Coordinatrice del Progetto Open Sport per Sport e Salute S.p.A., è anche redattrice per AIF Learning News, rivista digitale dell'Associazione Italiana Formatori, concentrata sull'integrazione dell'AI nei percorsi formativi.

**Massimo Conte** [conte@complexityeducation.com](mailto:conte@complexityeducation.com)

Digital Learning Innovation Manager presso E-CO e-learning Studio. Si occupa di progettazione didattica nell'e-learning da oltre 15 anni, e attualmente il suo focus è sulla ricerca e sviluppo per l'integrazione dell'AI generativa all'interno dei percorsi formativi digitali. È il Coordinatore Editoriale del Complexity Education Project, gruppo di ricerca dedicato allo studio e alla diffusione di metodi e conoscenze su reti e sistemi complessi



# Natural Language Processing e Computer Vision per l'accessibilità dei video di lezioni con formule

Tiziana Armano<sup>1</sup>, Gabriele Baratto<sup>3</sup> Cristian Bernareggi<sup>1</sup>, Sara Bertoldo<sup>4</sup>, Erika Brunetto<sup>1</sup>, Anna Capietto<sup>1</sup>, Manuela Caramagna<sup>3</sup>, Sandro Coriasco<sup>1</sup>, Nazareno Defrancesco<sup>4</sup>, Mattia Ducci<sup>1</sup>, Maria Luisa Gabrielli<sup>4</sup>, Mariafrancesca Guadalupi<sup>4</sup>, Alessandro Mazzei<sup>2</sup>, Angelo Saccà<sup>3</sup>, Francesco Tarasconi<sup>4</sup>

<sup>1</sup>Dipartimento di Matematica – Università di Torino, <sup>2</sup>Dipartimento di Informatica – Università di Torino, <sup>3</sup>Direzione SIPE - Università di Torino, <sup>4</sup>Maize srl

**Abstract.** Il contributo presenta il progetto VoiceMath (iniziato nel 2021) del Laboratorio Polin per la ricerca e la sperimentazione di nuove tecnologie assistive per le STEM, in collaborazione con Maize srl e la Direzione SIPE dell'Università di Torino. VoiceMath è un'applicazione web sviluppata per rispondere alle esigenze di accessibilità di video di lezioni universitarie di contenuto scientifico. Punti di forza del progetto sono: l'efficace trasformazione di un prodotto della ricerca in servizio per l'ateneo, l'impiego di tecnologie innovative per lo sviluppo di uno strumento complesso, l'evoluzione delle funzionalità in relazione ai nuovi sviluppi delle tecnologie utilizzate.

**Keywords.** LLM, NLP, computer vision, accessibilità, formule, speech-to-text

## Introduzione

Il periodo pandemico ha accelerato il processo di innovazione digitale nella didattica aumentando l'utilizzo di strumenti e risorse digitali ormai di uso comune. Tuttavia, l'utilizzo di strumenti e risorse digitali ha incrementato in alcuni casi il divario nei confronti di studenti con disabilità e DSA, che rischiano di essere esclusi dai benefici dell'innovazione. Una delle risorse didattiche digitali maggiormente utilizzata e diffusa è sicuramente il video della registrazione delle lezioni. Questa risorsa didattica digitale però presenta diversi problemi di accessibilità: persone con disabilità uditive non accedono all'audio, persone con disabilità visive non accedono ai contenuti video, persone con disabilità motorie e con DSA hanno difficoltà a prendere appunti. Per garantire l'accessibilità e la completa fruizione di tale risorsa servono trascrizione (Armano et al., 2020) e sottotitoli adeguati. Attualmente ci sono molti sistemi per la trascrizione e la sottotitolazione automatica, con buone performance e in continuo e veloce miglioramento: molti atenei forniscono sistemi e servizi per la trascrizione e la sottotitolazione anche in tempo reale. Tuttavia, nel caso di "lezioni con formule" la trascrizione del parlato in linguaggio naturale non risulta adeguata: risulta, infatti, ambigua, prolissa, non navigabile. È necessario che le formule siano codificate con un linguaggio opportuno, come LaTeX o MathML (Armano et al., 2021).

## 1. Il progetto VoiceMath

Lo sviluppo di uno strumento informatico per la trascrizione di “lezioni con formule” è stato possibile solo grazie all'evoluzione dei sistemi di speech-to-text e di riconoscimento ottico delle formule. Nel 2021, grazie a un finanziamento della Fondazione CRT e del Dipartimento di Matematica dell'Università di Torino, il Laboratorio Polin ha avviato il progetto Voice-Math con Maize srl (Amore et al, 2021). Il prototipo del software, disponibile a fine 2021, per trascrizioni integrate con LaTeX, è risultato promettente. Quindi la Direzione Sistemi Informativi Portale e E-learning dell'Università di Torino ha deciso di finanziarne l'integrazione con i servizi di ateneo insieme allo sviluppo di altre funzionalità. A ottobre 2022 l'applicativo è stato reso disponibile a un gruppo di docenti per una fase di test e, in seguito alle recenti evoluzioni nell'ambito dei Large Language Model, sono state implementate nuove funzionalità avanzate.

### 1.1 Caratteristiche tecniche

VoiceMath ha come obiettivo la produzione automatica di un flusso ordinato di contenuti in un documento LaTeX, a partire da (1) videoregistrazione di una lezione di discipline STEM e (2) corrispondenti immagini contenenti il testo scritto a mano dal docente. L'audio del docente, isolato dalla videoregistrazione, viene sottoposto a trascrizione automatica mediante tecnologie di Natural Language Processing (Jurafsky & Manning, 2012) e di Speech-to-Text.

Le immagini, solitamente ricavate dal pdf prodotto dal docente durante la videolezione, vengono sottoposte ad una componente di OCR ed Image-to-LaTeX, ovvero estrazione automatica delle formule contenute all'interno delle immagini, e trasformazione in linguaggio LaTeX.

Le trascrizioni dell'audio e delle immagini sono consultabili tramite un'applicazione web dove l'utente può rivedere/modificare le trascrizioni prodotte dai componenti automatici e combinarle per produrre un solo out-put in formato LaTeX. Il software fornisce alcuni strumenti che affiancano e aiutano l'utente in questa attività: modelli di classificazione e similarità fra testo e formule, player per la navigazione dell'audio estratto dalla videolezione, un ambiente di anteprima e modifica delle formule LaTeX, funzionalità di editing avanzate (Figura 1).

Fig. 1  
Screenshot  
dell'applicazione  
VoiceMath

The screenshot displays the VoiceMath application interface. On the left, a transcript of a lecture is visible, containing text such as: "Proviamo iniziare prima di iniziare la lezione vera e propria. Vorrei solo rievocare una cosa che riguarda quella piccola indagine a cui gentilmente avete risposto e a una persona che però non posso sapere se è in aula o no che ha fatto la laurea triennale in matematica finanziaria, non dalla facoltà che fate e lei. Ah scusi niente ha risposto quel quella mail in cui le ho. Comunque ho risposto. Insomma dicendo che non oggi ma dalla settimana prossima entreranno in gioco nei discorsi riguardanti un capitolo che si fa nella triennale in matematica che si chiama teoria della misura integrabile di 1900. Ah bene, insomma lo pensavo e lei ma in realtà a tutti noi ha scritto sulla pagina Moodle del corso una dispensa di questi argomenti. Sì." Below this, there are several numbered points (1-10) discussing mathematical topics like differential equations, linear operators, and vector spaces.

On the right, there is a section titled "Scegli il tag con cui vuoi copiare le formule:" with a dropdown menu set to "L<sub>1</sub>-L<sub>1</sub>". Below this, a list of detected formulas is shown, each with a copy icon and a delete icon. The formulas listed are:
 

- $f: X \rightarrow y \quad x_0 \in X$
- $f \in C^n(x)$
- $(k2)(df)(x_0)$
- $\exists U(x_0) \quad \exists V(f(x_0))$
- $f: U \rightarrow V$

 At the top right of the interface, there are buttons for "Preview", "Formula", and "Edit". At the bottom left, there is a "Vedi PDF" button.

Principali componenti del software sono i modelli di classificazione e di similarità delle formule, che sono stati aggiornati nel 2024 in relazione all'evoluzione recente delle tecnologie utilizzate (LLM in particolare). La prima componente ha l'obiettivo di identificare le porzioni di testo ad alto contenuto matematico all'interno della trascrizione automatica. L'identificazione si basa su un modello di classificazione binaria, addestrato su dati di dominio, che analizza il testo e classifica gli span come "formula/non formula". La soluzione proposta si basa su un fine-tuning di un modello sentence transformer multilingua, basato su BERT, utilizzando l'approccio SetFit, un framework per il few-shot fine-tuning. Il dataset è composto da 13 testi di trascrizioni di lezioni matematiche, trascritte e annotate manualmente assieme al Dipartimento di Matematica; per l'annotazione si è utilizzato il tag <inizio formula>/<fine formula> per identificare gli span di testo ad alto contenuto matematico. Il dataset così composto è stato preprocessato e diviso in training set, validation set e test set. Sono stati eseguiti due diversi fine-tuning: il primo attuando un bilanciamento del dataset tra le due classi, il secondo senza applicare il bilanciamento. La seconda componente ha l'obiettivo di aiutare l'utente nella revisione delle porzioni ad alto contenuto matematico, suggerendo la possibile formula con cui sostituire il testo evidenziato. L'identificazione si basa su un modello che calcola la distanza tra la porzione di testo e la formula trascritta in LaTeX. Le porzioni di testo classificate da un modello precedente vengono ri-analizzate e preprocessate: nella fase di preprocessing vengono rivisti e parsificati alcuni termini specifici del dominio (ad esempio, "meno" in < - >) per trasformare il testo da linguaggio naturale alla codifica LaTeX della specifica formula. Il modello sviluppato lavora sul testo parsificato e sulle formule trascritte in LaTeX, restituendo una metrica di somiglianza tra i due oggetti in input che permette, quindi, di collegare automaticamente porzioni di testo e formule. Queste due componenti consentono di mettere a disposizione degli utenti la nuova funzionalità di suggerer, che permette all'utente di individuare più facilmente le porzioni di testo da sostituire con codice LaTeX e di collegare la relativa formula convertita da immagine a LaTeX.

## **1.2 Dal progetto di ricerca al servizio di Ateneo**

In seguito all'analisi del prototipo realizzato nel 2021, la Direzione Sistemi Informativi, Portale, E-Learning ha ritenuto opportuno proporre VoiceMath come servizio di ateneo. Si è resa quindi necessaria la definizione di possibili scenari per rendere la soluzione scalabile e sostenibile nel tempo. Fondamentale in questo senso è stato lo sviluppo di funzionalità relative alla gestione multi-utente, il supporto per una coda di elaborazioni e l'integrazione con il sistema single sign-on di Ateneo. L'intera infrastruttura è stata inoltre potenziata per supportare un bacino di utenti più ampio rispetto a quello coinvolto nella fase di prototipo.

In questa fase del progetto è stata predisposta anche una filiera di supporto analoga a quella degli altri servizi online gestiti dalla Direzione SIPE, che prevede la predisposizione di manualistica, un canale di contatto e il censimento degli utenti abilitati all'applicativo, in modo da raggiungerli per comunicazioni di servizio (manutenzioni, disservizi, ecc.).

Il progetto è quindi diventato servizio di Ateneo a supporto dell'obiettivo del Piano Strate-

gico 2021-2026 rispetto all'inclusività (azione 1.2.3 Individuare soluzioni per supportare soggetti in situazioni particolari di disagio e offrire loro le migliori condizioni di studio/lavoro).

Un riferimento al servizio è stato anche incluso negli Obiettivi di Accessibilità (rif. art. 9, comma 7, del D.L. 18 ottobre 2012, n. 179) per l'Ateneo per l'anno 2022.

## Riferimenti bibliografici

Amore M., Armano T., Bernareggi C., Capietto A., Coriasco S., Crespan R., Duc-ci M., Gabrielli M. L., Guadalupi M. F., Mazzei A., Mazzei A., Sofia A. Tarasconi F. (2021), Riconoscimento vocale di formule, Conferenza Garr 2021.

Armano T., Baratto G., Bernareggi C., Bosco M., Capietto A., Caramagna M., Ducci M., Mazzei A., Mazzei A., Tarasconi F. (2021), Matematica a voce parte 1: trascrizione di lezioni di matematica, Atti del MoodleMoot Italia 2020, (2021), pp 23-27.

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

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

## Autori

**Tiziana Armano** [tiziana.armano@unito.it](mailto:tiziana.armano@unito.it)

Tecnico della ricerca presso il Dipartimento di Matematica "G. Peano" dell'Università degli Studi di Torino. Si occupa di e-learning e tecnologie per la didattica. Ha fatto parte del gruppo di ricerca del progetto "Per una matematica accessibile e inclusiva" dal 2012 ed è membro del Laboratorio Polin. In questo ambito segue in particolare i progetti VoiceMath, SpeechMatE e Audiofunctions.web.

**Gabriele Baratto** [gabriele.baratto@unito.it](mailto:gabriele.baratto@unito.it)

Attualmente membro dello staff E-learning, Collaboration online e Architetture dell'Università di Torino. Sempre presso l'Ateneo torinese ha conseguito la laurea specialistica in Scienze Linguistiche, con una tesi in sociolinguistica, e il dottorato di ricerca in Scienze del Linguaggio e della Comunicazione, con una tesi in linguistica italiana. È stato collaboratore dell'Atlante Linguistico ed Etnografico del Piemonte Occidentale (ALEPO), docente a contratto presso l'Università di Torino, docente di scuola secondaria e coordinatore di diversi progetti per contrastare l'insuccesso scolastico.

**Cristian Bernareggi** [cristian.bernareggi@unimi.it](mailto:cristian.bernareggi@unimi.it)

Dottorato di ricerca in informatica nel 2006. Tra il 2004 e il 2009 ha condotto ricerca in ambito HCI nei progetti europei LAMBDA ed @Science. Dal 2023 è tecnico della ricerca presso il Dipartimento di Matematica "G. Peano" dell'Università degli Studi di Torino. Tra il 2011 e il 2018 ha partecipato alla fondazione e alla crescita dello spin-off EveryWare Technologies per lo sviluppo di tecnologie assistive su dispositivi mobili. Dal 2018 collabora con il Laboratorio Polin sull'accessibilità alle STEM per persone con disabilità.

**Sara Bertoldo** [sara.bertoldo@maize.io](mailto:sara.bertoldo@maize.io)

Lavora come linguista computazionale e sviluppatrice in Maize. Laureata in Lingue, civiltà e scienze del linguaggio presso l'università Ca' Foscari di Venezia e successivamente in Informatica Umanistica presso l'università di Pisa. Coniuga queste due anime in progetti di analisi di testo, assistenti conversazionali, generazione e processamento del linguaggio naturale dal forte carattere consulenziale.

**Erika Brunetto** [erika.brunetto@unito.it](mailto:erika.brunetto@unito.it)

Laureata magistrale in Matematica presso l'Università degli Studi di Genova. Dopo la laurea ha avuto diverse esperienze sia in ambito didattico che aziendale. Dopo un assegno di ricerca di due anni presso CNR-IMATI (Istituto di Matematica Applicata e Tecnologie Informatiche) della sede di Genova, da più di un anno collabora con il Laboratorio Polin di Unito come borsista di ricerca.

**Anna Capietto** [anna.capietto@unito.it](mailto:anna.capietto@unito.it)

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

**Manuela Caramagna** [manuela.caramagna@unito.it](mailto:manuela.caramagna@unito.it)

Responsabile dello Staff E-Learning, Collaboration on line, Architetture presso la Direzione Sistemi Informativi, Portale, E-learning dell'Università di Torino. Laureata in Comunicazione Multimediale e di Massa e con un Master in IT Governance and Compliance, si occupa dal 2006 di gestione di servizi web e di tecnologie per la didattica.

**Sandro Coriasco** [sandro.coriasco@unito.it](mailto:sandro.coriasco@unito.it)

Professore Associato di Analisi Matematica presso il Dipartimento di Matematica dell'Università di Torino. Referente per l'inclusione di studenti e studentesse con DSA. Si occupa di analisi di Fourier e di analisi globale su varietà non compatte e su varietà con singolarità. È un componente del gruppo di ricerca sulle tecnologie assistive presso il Laboratorio "S. Polin". In tale ambito si occupa, in particolare, dello sviluppo del pacchetto LaTeX accessibility e, più in generale, dei temi collegati all'accessibilità dei contenuti matematici.

**Nazareno Defrancesco** [nazareno.defrancesco@maize.io](mailto:nazareno.defrancesco@maize.io)

Deep Learning e Software Analyst presso Maize srl.

**Mattia Ducci** [mattia.ducci@unito.it](mailto:mattia.ducci@unito.it)

Laurea magistrale in informatica all'Università degli Studi di Milano nel 2019. Durante e dopo l'università ha collaborato a vari progetti incentrati sul tema dell'accessibilità: da soluzioni su dispositivi mobili per persone con disabilità visive (Università degli Studi di Milano, Università degli Studi di Pavia) a soluzioni per rendere più accessibili le materie scientifiche (Università degli Studi di Torino). Ha lavorato dal 2018 al 2023 come consulente informatico in un'azienda per sviluppo mobile iOS, cross-platform e full-stack e ora lavora come libero professionista su temi legati

all'accessibilità.

**Maria Luisa Gabrielli** [marialuisa.gabrielli@maize.io](mailto:marialuisa.gabrielli@maize.io)

Lavora come Data Scientist presso Maize. Laureata in Matematica presso l'Università di Torino, si è specializzata in Analisi numerica, in particolare avvicinandosi al mondo dell'elaborazione e del riconoscimento di immagini. Ha esperienza in Data Analysis, nell'applicazione di algoritmi di Machine Learning e Deep Learning e in tecniche di Computer Vision.

**Mariafrancesca Guadalupi** [mariafrancesca.guadalupi@maize.io](mailto:mariafrancesca.guadalupi@maize.io)

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

**Alessandro Mazzei** [alessandro.mazzei@unito.it](mailto:alessandro.mazzei@unito.it)

Laurea in Fisica (indirizzo Cibernetico) presso l'Università Federico II di Napoli nel 2000. Dottore di Ricerca in Informatica presso l'Università degli Studi di Torino nel 2006, dove è Professore Associato presso il Dipartimento di Informatica. I suoi interessi di ricerca includono le ontologie giuridiche, la sintassi e la semantica computazionale, la generazione automatica del linguaggio, i sistemi di dialogo, il ragionamento automatico, le tecnologie assistive e il food computing. È autore di oltre 100 lavori scientifici peer reviewed.

**Angelo Sacca** [angelo.sacca@unito.it](mailto:angelo.sacca@unito.it)

Direttore della Direzione Sistemi Informativi, Portale, E-learning dell'Università di Torino. Presidente dell'Associazione Italiana Comunicatori d'Università (AICUN).

**Francesco Tarasconi** [francesco.tarasconi@maize.io](mailto:francesco.tarasconi@maize.io)

Data Scientist and Artificial Intelligence Manager in MAIZE. In MAIZE si occupa di Natural Language Processing, AI generativa e sistemi di supporto alle decisioni. Ha supportato ENI nella realizzazione del Progetto di AI Safety Pre-Sense per la tutela dell'integrità fisica dei dipendenti. Collabora con l'Università di Torino nello sviluppo di software con l'obiettivo di migliorare l'accessibilità alle lezioni di materie STEM. Conduce attività di ricerca industriale in collaborazione con il Politecnico di Torino su temi di AI applicata all'ambito giuridico.

# Artificial Intelligence, LLM and RAG supporting Library and Information Science: challenges and opportunities

Giorgia Di Marcantonio

University of Macerata (Italy)

**Abstract.** ICTs are transforming the methods Library and Information Science uses to organize and disseminate knowledge. This transformation is evident in the adoption of new descriptive standards (RDA/RIC-CM) and numerous research projects (such as INTERPARES TRUST AI) that seek to assess and manage the impact of technologies, especially artificial intelligence, on archival science and librarianship within a global framework. The contribution offers insights into the future of knowledge description and cataloging, emphasizing the enhanced access to informational resources enabled by the use of LLM and RAG in digital environments designed for exploring cultural heritage

**Keywords.** Artificial Intelligence; Archives; Library and Information Science; LLMs; RAG;

The critical situation that led to the National Recovery and Resilience Plan (PNRR)<sup>1</sup> has secured substantial funding for Italy's "digital transition". The Plan dedicates significant financial resources to "Digitalization, innovation, competitiveness, culture, and tourism," amounting to €40.29 billion (Di Marcantonio and Valacchi 2022). This investment includes a major focus on cultural heritage, aiming to establish a vast repository of digitized resources to be housed in a National Digital Library (Cerullo and Negri 2023).

Researchers seeking information on Italy's cultural heritage online already have access to numerous digital platforms and information systems. These include the National Archival System (SAN)<sup>2</sup>, the Open Public Access Catalogue (OPAC-SBN)<sup>3</sup>, Internet Culturale<sup>4</sup>, Alfabetica<sup>5</sup>, the Information System of the State Archives<sup>6</sup>, Digital Archive<sup>7</sup>, Manus Online<sup>8</sup>, Cultura Italia<sup>9</sup>, and the General Catalog of Cultural Heritage<sup>10</sup>, among others. Each platform has its own navigation paths, and while some offer direct access to digitized cultural resources, others provide only summary information, indicating whether a resource exists, where it is stored, and how it can be accessed within cultural institutions. In Italy, there is a lack of comprehensive studies on user satisfaction with these platforms (Feliciati 2016). The few studies that do exist provide limited insight, leaving a significant gap in understanding how effectively these resources serve their users (Craig 1998; Chapman 2010). Navigating these platforms can be especially difficult for non-expert users due to the prevalent use of technical language that can be hard to understand for those not well-versed in the field. This complexity not only restricts access to valuable cultural information, but also deters wider public engagement. Simplifying user interfaces and providing more intuitive navigation and search capabilities could greatly improve the user experience and make these resources more accessible (Djamasbi, Siegel, and Tullis 2010).

Digital platforms frequently lack the cultural mediation between scholars and resources that is typically present in archives, libraries, and museums (Michetti 2020). This highlights the necessity of thoughtfully presenting and organizing cultural knowledge online. The target group of users of these platforms is varied, encompassing students, scholars, researchers, and ordinary citizens who deserve easy access to their cultural heritage. However, creating navigation paths that satisfy all potential users of digital cultural resources is a significant challenge. The resources are varied, different description standards are employed, and each scientific field has its unique requirements that it is reluctant to compromise. Enhancing access systems could greatly benefit from the use of Artificial Intelligence (AI). Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) could facilitate natural interactions between users and platforms (Kasneci et al. 2023).

Large Language Models (LLMs) are a type of AI that can understand and generate human language. Trained on vast amounts of text data, they can perform various language tasks, such as translation, summarization, and question-answering. Models like OpenAI's GPT-4 can understand context, infer meaning, and generate coherent and contextually relevant responses, making them powerful tools for enhancing user interaction with digital platforms.

The integration of Artificial Intelligence (AI) in managing cultural resources presents both exciting opportunities and significant challenges. AI systems, especially Large Language Models (LLMs), have the potential to produce inaccurate results, known as hallucinations, which can stem from biases in the training data (Ji et al. 2023). Additionally, AI systems can perpetuate stereotypes and manipulate information, raising ethical concerns. In the context of cultural resource management, these limitations are particularly problematic because the goal is to provide reliable, scientifically validated information for research and self-education.

One major issue with LLMs is their tendency to generate plausible-sounding but incorrect or misleading information. This is especially critical when dealing with cultural heritage, where accuracy and reliability are paramount. Users need to trust that the information they access about cultural artifacts, historical documents, or scholarly works is both accurate and authoritative.

To mitigate the issues inherent in LLMs, a promising approach is combining them with Retrieval-Augmented Generation (RAG) systems (Lewis et al. 2020). RAG systems enhance the relevance and accuracy of generated responses by integrating a retrieval mechanism that draws upon a pre-defined knowledge base. This dual approach ensures that responses are grounded in verifiable data.

In practical terms, a RAG system starts by retrieving relevant documents or data from a curated database or a selected set of resources based on the user's query. This retrieved information then guides the text generation process, ensuring that the responses are contextually appropriate and factually accurate.

A significant advantage of RAG systems is their ability to address and mitigate biases inherent in training data. By relying on carefully selected databases, RAG systems ensure

the information used to generate accurate and free from distortions responses. This is crucial in cultural resource management, where information integrity is paramount.

To further enhance reliability, these selected databases are often converted into a vectorized format. This process involves converting texts into numerical sequences using embedding algorithms, enabling the system to perform comparisons based on vector similarity. By comparing the input vector with those representing information in the database, the system can identify the most relevant resources to answer the user's query. Techniques such as cosine similarity are commonly used to ensure the most pertinent information is retrieved and utilized.

As AI continues to be integrated into cultural resource management, it is essential to address the ethical implications. The capacity of AI to influence user opinions through language manipulation is a double-edged sword. While it can make interactions more engaging and informative, it also has the potential to mislead if not carefully managed (Chen et al. 2024). Ensuring transparency in AI operations and maintaining rigorous standards for data accuracy and bias mitigation are critical for fostering trust and reliability.

Continuous user feedback and systematic studies on user satisfaction and accessibility are necessary to refine these digital platforms. Understanding the diverse needs of users, from scholars to everyday citizens, will aid in designing more intuitive and inclusive interfaces, making cultural heritage more accessible to all.

LLMs, whether used alone or with RAG integration, do not overhaul the existing systems used for managing cultural resources. Catalog entries, archival descriptions, and inventories should—and must—remain accessible to users. What these technologies enhance are the points of access to information, facilitating that initial exploration or navigation when it's unclear what resources are available on a given topic. RAG systems increase the likelihood of serendipitous discoveries and provide more structured responses to user queries.

Incorporating these technologies into knowledge management environments can significantly reduce the time needed to find and utilize resources. However, they should be viewed as supplementary features rather than replacements for the existing systems. By adding these advanced functionalities, users can benefit from faster, more intuitive access to the wealth of information contained within traditional cataloging and archival systems, without losing the depth and reliability these systems provide (Di Marcantonio 2024).

Integrating LLMs and RAG systems could greatly enhance access to cultural resources on digital platforms. However, it is crucial to invest in the training of industry professionals, data modeling, and the technological tools available to cultural institutions. The digital transformation requires significant investments, which have been made possible in Italy through the PNRR. However, embracing digital methodologies is equally crucial. To simply have the tools is insufficient; a "digital mindset" is essential to produce user-friendly and accurate contents that are potentially available in long-term for future generations—the primary target of the NEXGenerationEU funds.

## References

Cerullo, Luigi, e Antonella Negri. 2023. «L'infrastruttura software per il patrimonio culturale (ISPC) come abilitatore di un Ecosistema digitale nazionale del patrimonio culturale». *Digitalia* 18 (1): 38–50. <https://doi.org/10.36181/digitalia-00059>.

Chapman, Joyce Celeste. 2010. «Observing Users: An Empirical Analysis of User Interaction with Online Finding Aids». *Journal of Archival Organization* 8 (1): 4–30. <https://doi.org/10.1080/15332748.2010.484361>.

Chen, Xiaoyang, Ben He, Hongyu Lin, Xianpei Han, Tianshu Wang, Boxi Cao, Le Sun, e Yingfei Sun. 2024. «Spiral of Silences: How is Large Language Model Killing Information Retrieval? A Case Study on Open Domain Question Answering». <https://doi.org/10.48550/ARXIV.2404.10496>.

Craig, Barbara Lazenby. 1998. «Old Myths in New Clothes: Expectations of Archives Users». *Archivaria* 45 (February): 118–26. <https://archivaria.ca/index.php/archivaria/article/view/12228>.

Di Marcantonio, Giorgia. 2024. «Artificial Intelligence, Large Language Models (LLMs), and Retrieval-Augmented Generation (RAG). New Tools for Accessing Archival and Bibliographic Resources». *Bibliothecae.it* 13 (1).

Di Marcantonio, Giorgia, e Federico Valacchi. 2022. «Tra dematerializzazione e next generation: i sistemi documentari nel PNRR». *Rivista elettronica di diritto, economia, management* 2:16–32.

[https://www.clioedu.it/documenti/eventi-live-ondemand/rivista-elettronica/Rivista-elettronica-2\\_2022.pdf](https://www.clioedu.it/documenti/eventi-live-ondemand/rivista-elettronica/Rivista-elettronica-2_2022.pdf)

Djamasbi, Soussan, Marisa Siegel, e Tom Tullis. 2010. «Generation Y, Web Design, and Eye Tracking». *International Journal of Human-Computer Studies* 68 (5): 307–23. <https://doi.org/10.1016/j.ijhcs.2009.12.006>.

Feliciati, Pierluigi. 2016. «L'usabilità degli ambienti bibliotecari e archivistici digitali come requisito di qualità: contesto, modelli e strumenti di valutazione». *JLIS.it* 7 (1): 113–30. <https://doi.org/10.4403/jlis.it-11512>

Ji, Ziwei, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, e Pascale Fung. 2023. «Survey of Hallucination in Natural Language Generation». *ACM Computing Surveys* 55 (12): 1–38. <https://doi.org/10.1145/3571730>.

Kasneci, Enkelejda, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementeva, Frank Fischer, Urs Gasser, et al. 2023. «ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education». *Learning and Individual Differen-*

ces 103. <https://doi.org/10.1016/j.lindif.2023.102274>.

Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, et al. 2020. «Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks». [Preprint] <https://doi.org/10.48550/ARXIV.2005.11401>.

Michetti, Giovanni. 2020. «Il mondo come puzzle. I beni culturali nel web». *Digitalia Web Rivista del digitale nei beni culturali* 15 (1): 26–42. <https://doi.org/10.36181/digitalia-00002>

## Notes

<sup>1</sup><https://www.governo.it/sites/governo.it/files/PNRR.pdf> (Last access: 6/24/24).

<sup>2</sup><http://san.beniculturali.it> (Last access: 6/24/24).

<sup>3</sup><https://opac.sbn.it> (Last access: 6/24/24).

<sup>4</sup><https://www.internetculturale.it/> (Last access: 6/24/24).

<sup>5</sup><https://alphabetica.it/web/alphabetica/> (Last access: 6/24/24).

<sup>6</sup><https://sias.archivi.beniculturali.it/> (Last access: 6/24/24).

<sup>7</sup><https://www.archiviodigitale.icar.beniculturali.it> (Last access: 6/24/24).

<sup>8</sup><https://manus.iccu.sbn.it/> (Last access: 6/24/24).

<sup>9</sup><https://www.culturaitalia.it/> (Last access: 6/24/24).

<sup>10</sup><https://catalogo.beniculturali.it> (Last access: 6/24/24).



## Author

Giorgia Di Marcantonio [g.dimarcantonio@unimc.it](mailto:g.dimarcantonio@unimc.it)

Giorgia Di Marcantonio is a researcher in Archival and Library Science, (SSD M-5T0/08) at the University of Macerata. Her main research areas explore the relationship between archives and ICTs, with a special focus on international standards and models for information retrieval. She is a member of the scientific committee for the Italian Journal of Library, Archives and Information Science (JLIS.it) and has published numerous articles in sector-specific scientific journals. Recently, she has been involved with the international project InterPARES Trust AI (2021-2026), which aims to delineate the risks and benefits associated with the use of artificial intelligence in record management.

# Il patrimonio culturale come sistema di relazioni. Verso un nuovo Sistema del Catalogo Nazionale

Carlo Birrozzi, Fabrizio Magnani, Elena Musumeci, Chiara Veninata

ICCD – Istituto Centrale per il Catalogo e la Documentazione

**Abstract.** Il contributo presenta gli esiti di un profondo ripensamento dei processi, delle metodologie e delle tecnologie a supporto della catalogazione dei beni culturali, nell'ambito della progettazione di un nuovo Sistema del Catalogo Nazionale e nel solco delle iniziative intraprese dall'ICCD nel campo del web semantico. Dalla creazione di un knowledge graph del patrimonio culturale italiano e di una rete di ontologie come riferimento per la rappresentazione delle informazioni sui beni e sulle entità di interesse culturale, si giunge all'ideazione di un nuovo sistema che vuole razionalizzare i flussi di produzione dei dati sfruttando i linked open data in fase di generazione del dato catalografico. Lo sviluppo dell'applicativo, tuttora in corso, si fonda sull'analisi delle attuali criticità, in un approccio iterativo e incrementale e su un importante cambiamento di paradigma che risponde all'esigenza di restituire i beni culturali al loro contesto, in una complessità di relazioni

**Keywords.** Catalogazione, beni culturali, patrimonio culturale, ontologie, linked open data

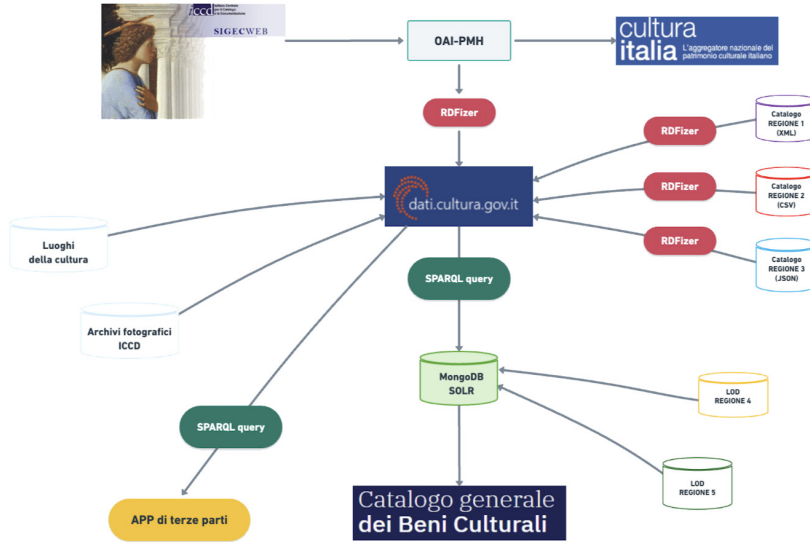
In seno al Ministero della Cultura, l'ICCD gestisce il Catalogo generale dei beni culturali, fonte primaria di censimento e conoscenza scientifica di tutto il patrimonio culturale, ad esclusione dei beni archivistici e librari. L'Istituto è responsabile della definizione di standard e metodologie di catalogazione, secondo criteri omogenei a livello nazionale. Negli ultimi anni ha investito fortemente nelle tecnologie semantiche pubblicando l'intero database del catalogo in linked open data (LOD) sulla base delle ontologie di ArCo - Architettura della conoscenza, il grafo della conoscenza del patrimonio culturale italiano, realizzato e ampliato in collaborazione con l'ISTC del CNR e con l'Università di Bologna (Birrozzi et al. 2020; Veninata 2020; Carriero et al. 2019a, 2019b, 2019c)<sup>1</sup>. A marzo 2021, è stato rilasciato il nuovo sito di consultazione dei dati del Catalogo basato su tali tecnologie<sup>2</sup>.

Il SiGECweb è il Sistema Informativo Generale del Catalogo che l'ICCD mette a disposizione degli Istituti del MiC, e di altri soggetti coinvolti nelle attività di catalogazione: Regioni, Comuni, Università, enti ecclesiastici e ogni altro soggetto abilitato come ente schedatore. Il sistema integra e recupera anche risorse prodotte da sistemi diversi, che possono conferire dati nel rispetto degli standard ministeriali o sulla base di possibili mappature. I dati presenti nel SiGECweb sono convertiti automaticamente in LOD e resi accessibili attraverso il sito ministeriale dei dati aperti (<https://dati.cultura.gov.it>). Il sito funge principalmente da interfaccia machine-to-machine, offrendo LOD interrogabili liberamente da chiunque, tramite uno SPARQL endpoint<sup>3</sup>.

Possono qui confluire dati che le Regioni pubblicano in open data – che l'ICCD converte in

LOD secondo le ontologie di ArCo<sup>4</sup> – o i dati pubblicati direttamente come LOD, che possono essere così recuperati e pubblicati sul sito di consultazione del Catalogo. L'obiettivo è quello di favorire flussi di interoperabilità tra dati, per superare una logica “a silos”, con banche dati chiuse e non dialoganti, verso la costituzione di un catalogo nazionale.

Fig. 1  
Il flusso dei dati:  
lo stato attuale



Nel maggio 2023, ha preso avvio lo sviluppo di un nuovo Sistema del Catalogo Nazionale (SCN), a superamento dell'attuale SIGECweb<sup>5</sup>.

Il nuovo Sistema informativo è stato ribattezzato CLIO, riconoscendo nell'acronimo il ruolo fondazionale delle attività di “catalogazione”, “localizzazione”, “Identificazione” e “organizzazione” della conoscenza dei beni culturali. Il richiamo a Clio, musa della Storia, rimanda a un catalogo che contribuisce a preservare la memoria e il valore del patrimonio culturale, ispirandone la narrazione complessa.

La progettazione del nuovo sistema ha coinciso con una riflessione che rinnova il senso e le pratiche del Catalogo: un ripensamento profondo delle procedure, degli strumenti e delle metodologie di documentazione, descrizione e restituzione dei beni culturali di diversa tipologia. Il progetto si fonda sull'analisi delle attuali criticità, segue una metodologia agile e un approccio iterativo e incrementale<sup>6</sup>. “Navigare la complessità” è il titolo di questo evento del GARR ed è anche la finalità principale che guida lo sviluppo di CLIO, in un significativo cambiamento di paradigma. La descrizione del patrimonio culturale necessita, infatti, di una ridefinizione di strumenti e metodi volti a ricomporre in un sistema di relazioni complesse dati e risorse digitali disperse. Il nuovo sistema intende razionalizzare i flussi di produzione dei dati sfruttando i LOD già in fase di generazione del dato catalogafico (catalinking)<sup>7</sup>.

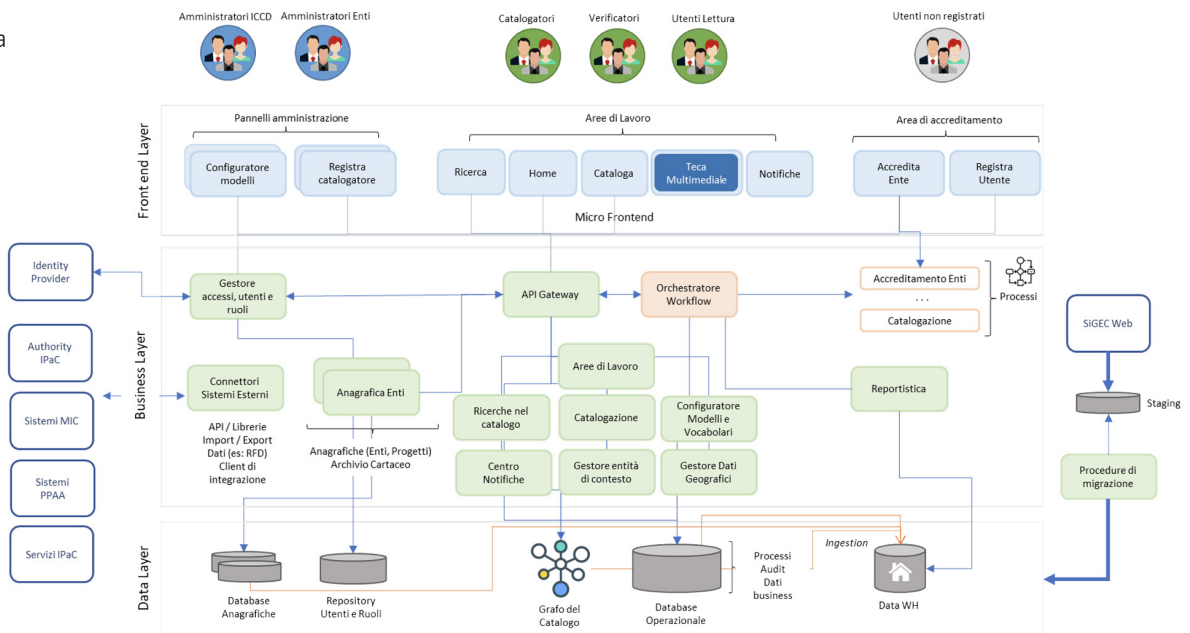
Uno dei requisiti fondamentali del nuovo sistema riguarda la “qualità del dato”, affidato

alla competenza dei catalogatori, ma ora costantemente arricchito e contestualizzato, aumentando dinamicamente il knowledge graph del patrimonio culturale. Ciò avviene tramite strumenti che vanno dall'interlinking all'accesso a fonti autorevoli e sistemi esterni per acquisire dati e risorse o stabilire collegamenti (OPAC-SBN, Iconclass, VIAF, Wikidata, ecc.), dall'uso di vocabolari controllati alla semplificazione delle attività di revisione dei dati, fino all'inclusione di diverse tipologie di risorse multimediali, annotabili semanticamente. Tale scenario prevede anche l'integrazione di servizi digitali avanzati (intelligenza artificiale, authority file nazionale, teca multimediale etc.) offerti da I.PaC, Infrastruttura e servizi digitali per il Patrimonio Culturale, il cui sviluppo rientra tra le azioni delineate nel Piano nazionale di digitalizzazione del patrimonio culturale<sup>8</sup>, elaborato dalla Digital Library del Ministero della cultura<sup>9</sup>.

Attraverso il nuovo sistema del Catalogo, si cerca di realizzare il passaggio da un modello oggetto-centrico, ereditato dalle prassi catalografiche affermatesi storicamente, a un modello che pone al centro il contesto di riferimento dei beni culturali, al fine di favorire una lettura del patrimonio come sistema di relazioni. Si procede così dal generale al particolare, dal contesto alla specifica entità culturale descritta.

L'organizzazione a grafo supporta questo cambiamento, superando il tradizionale approccio alla descrizione di singoli beni, con dati e documenti chiusi e conchiusi in una scheda di catalogo, in favore dell'arricchimento dinamico di dati, risorse e relazioni. Attraverso i LOD, possiamo navigare e interrogare un mondo aperto di connessioni potenzialmente infinite.

Fig. 2  
CLIO: Architettura logica

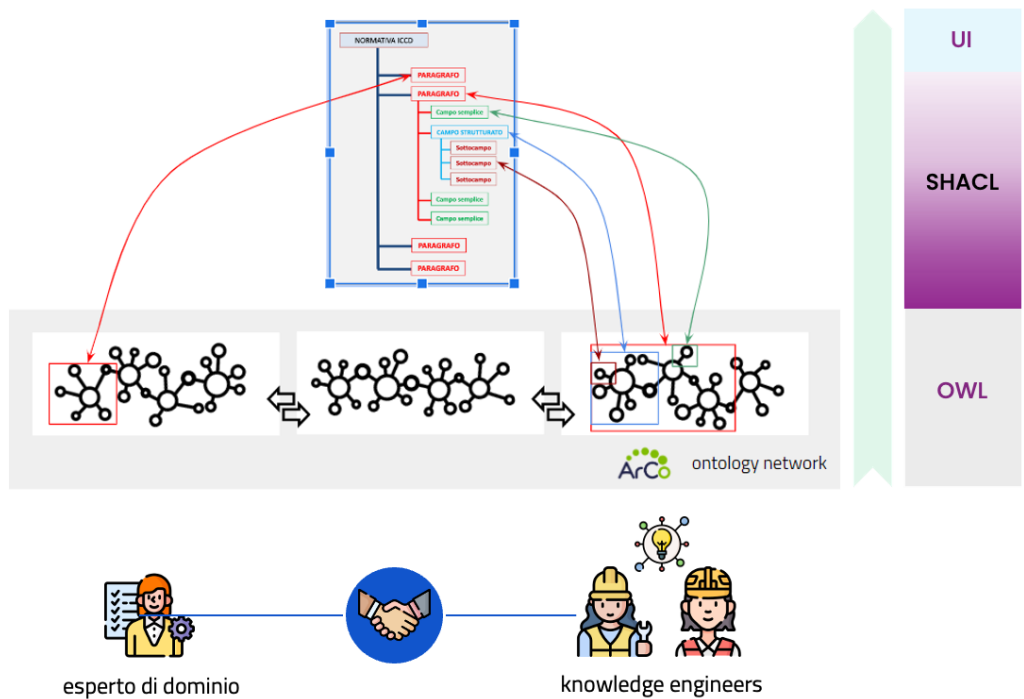


L'architettura logica di CLIO è stata progettata per essere adattabile e integrabile con

servizi e infrastrutture cloud, modulare e altamente scalabile. Il sistema si basa su una struttura a micro-servizi, articolata fondamentalmente su tre livelli: le interfacce grafiche, che includono diverse aree di lavoro personalizzabili, per consentire all'utente di svolgere attività simultanee e parallele; i servizi, tra i quali troviamo un Configuratore di Modelli e Vocabolari che gestirà il ciclo di vita dei modelli ontologici<sup>10</sup>; uno strato di persistenza, che supporta la modellazione di informazioni eterogenee, come geodata, risorse e dati descrittivi sui beni, dati relativi all'operatività degli utenti. Questo strato include un database a grafo che organizzerà semanticamente le relazioni tra tutte le entità.

Si passa da un processo attuale, in cui i LOD vengono generati a valle della catalogazione e all'esterno del Sistema del catalogo, secondo il modello derivato delle ontologie di ArCo<sup>11</sup>, a un modello che ne prevede la produzione nativa, all'interno di CLIO. In questo nuovo approccio, esperti di dominio ed esperti di modellazione governano insieme l'evoluzione e il ciclo di vita dei modelli per la rappresentazione di beni/entità di interesse culturale e di altre entità di contesto, gestendo regole e vincoli tramite SHACL<sup>12</sup>. In una diversa concezione degli standard ministeriali, i modelli descrittivi non si attesteranno più su interi tracciati predefiniti al momento del loro rilascio, ma su nuclei informativi formalizzati più ristretti che potranno evolvere in versioni successive.

Fig. 3  
Modelli di catalogazione (to be)



Tutte attività di catalogazione si svolgono all'interno di progetti per la conoscenza, documentazione e descrizione dei beni culturali. Il "Progetto" assume un'importanza decisa all'interno del nuovo sistema: diventa il tramite attraverso cui si dispiega il flusso di

catalogazione e l'applicazione dei diversi profili di visibilità e operatività sul dato. Ogni progetto prevede la definizione di uno o più ambiti di interesse disciplinare (archeologico, architettonico, etnoantropologico, ecc.) per qualificare e caratterizzare le attività di catalogazione, determinando quali moduli informativi, eventualmente specializzati, siano a disposizione dei catalogatori. Sarà inoltre possibile lavorare in modo collaborativo alla catalogazione, promuovendo una maggiore condivisione interdisciplinare delle conoscenze e migliorando l'accuratezza dei dati. Questa funzionalità mira a valorizzare una pluralità di competenze in grado di concorrere alla documentazione e alla descrizione dei beni e di altre entità culturali. L'ICCD ha definito oltre trenta tracciati catalografici, in relazione alle differenti tipologie di beni. Sebbene questa articolazione rifletta i settori disciplinari accademici, è ormai matura la consapevolezza di quanto i confini, nell'ambito dei beni culturali, siano difficili da stabilire e possano mutare nel tempo. Pertanto, attraverso il nuovo approccio modulare si vuole garantire la possibilità di espandere la descrizione dei beni con approfondimenti tecnico-specialistici interdisciplinari, a partire da un set minimo di dati, finalizzati alla loro identificazione univoca. Allo stesso modo, verranno introdotti strumenti di crowdsourcing, che consentiranno agli esperti di segnalare eventuali errori e integrare dati già pubblicati, mediante note e commenti altamente qualificati.

Lo sviluppo di CLIO si basa, infine, sull'assunto metodologico di dare piena centralità alla dimensione geografica dei patrimoni culturali e al loro stretto legame con i territori. Il nuovo sistema prevede numerosi scenari entro cui diventa possibile creare, gestire, richiamare il dato geografico: dalla gestione degli ambiti di operatività degli utenti e degli enti schedatori, alla gestione delle diverse localizzazioni qualificate di un bene culturale, combinando così la dimensione spaziale con quella temporale (localizzazione corrente, luogo di reperimento, luogo di origine etc.).

Il nuovo approccio segna un passaggio significativo, con importanti ricadute sulle metodologie di rilevamento, documentazione, catalogazione dei beni culturali, oltre che sulla loro restituzione su base territoriale, alimentando le molte potenziali applicazioni del dato geografico (conoscenza del territorio, tutela e pianificazione, gestione delle emergenze, ecc.). Soprattutto, il passaggio da un elenco dei beni a una mappa che rappresenta geograficamente le forme, i beni, le attività di un territorio, permette di "vedere le connessioni": una sorta di rappresentazione perspicua del patrimonio (Wittgenstein 1967, §122), in cui reinterpretare la densità della nozione stessa di patrimonio culturale e il senso stesso della sua catalogazione.

### **Riferimenti bibliografici**

C. Birrozzi, B. Barbaro, M.L. Mancinelli, A. Negri, E. Plances, C. Veninata (2020), Catalogare nel 2020, Aedon. Rivista di Arti e Diritto on line, Fasc. 3, 2020, pp. 191-196, Società Editrice il Mulino, Bologna,  
<http://www.aedon.mulino.it/archivio/2020/3/birrozzi.htm>

C. Veninata (2020), Dal Catalogo generale dei beni culturali al knowledge graph del patrimonio culturale italiano: il progetto ArCo, Digitalia. Rivista del digitale nei beni culturali,

15, 2, 2020, pp. 43-56, Istituto Centrale per il Catalogo Unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU).

V.A. Carriero, A. Gangemi, M.L. Mancinelli, A.G. Nuzzolese, V. Presutti, C. Veninata (2019a), Pattern-based design applied to cultural heritage knowledge graphs, 2019, ArXiv, abs/1911.07585.

V.A. Carriero, A. Gangemi, M.L. Mancinelli, L. Marinucci, A.G. Nuzzolese, V. Presutti, C. Veninata (2019b), ArCo: the Italian Cultural Heritage Knowledge Graph, 2019, ArXiv:1905.02840.

V.A. Carriero, A. Gangemi, M.L. Mancinelli, L. Marinucci, A.G. Nuzzolese, V. Presutti, C. Veninata (2019c), ArCo Ontology Network and LOD on Italian Cultural Heritage, A. Poggi (a cura di) Proceedings of the First International Workshop on Open Data and Ontologies for Cultural Heritage co-located with the 31st International Conference on Advanced Information Systems Engineering (CAiSE 2019, Rome, Italy, June 3, 2019), pp. 97-102. <http://ceur-ws.org/Vol-2375/>

Wittgenstein L. (1967), ed. or. 1953, Ricerche filosofiche, Einaudi, Torino.

## Autori



**Carlo Birrozzi** è Direttore dell'Istituto Centrale per il Catalogo e la Documentazione (ICCD) del Ministero della Cultura. È stato funzionario architetto presso la Soprintendenza per i beni architettonici ed il paesaggio di Milano. Ha lavorato presso la Direzione Generale per l'architettura e l'arte contemporanee e l'Istituto Centrale per il Restauro. È stato Soprintendente per i beni architettonici del Molise, delle Marche e della Città di Bari. Ha ricoperto il ruolo di Direttore ad interim dell'Istituto Centrale per i Beni Sonori e Audiovisivi (ICBSA). È curatore di mostre e autore di diverse pubblicazioni.

**Fabrizio Magnani**, etnoantropologo, ha condotto attività di ricerca in Italia e all'estero nel campo dell'espressività di tradizione orale, dell'antropologia della memoria e dell'antropologia visuale. Dal 2017 è funzionario demioetnoantropologo del Ministero della cultura. Lavora presso l'Istituto Centrale per il Catalogo e la Documentazione (ICCD), dove è responsabile del Servizio Gestione del sistema di catalogazione, pubblicazione, condivisione e valorizzazione dei dati e direttore dell'esecuzione per la realizzazione del nuovo Sistema del Catalogo Nazionale (CLIO-SCN).



**Elena Musumeci**, etnoantropologa, è specializzata in Beni archivistici e librari e dottore di ricerca in Scienze del libro e del documento presso l'Università La Sapienza. Attualmente lavora come funzionario demioetnoantropologo presso l'Istituto Centrale per il Catalogo e la Documentazione del Ministero dei Beni Culturali, dove è referente per la catalogazione del patrimonio demioetnoantropologico e responsabile del servizio Progetti e processi di catalogazione.



**Chiara Veninata** è direttrice dell'Istituto centrale per gli archivi. Fino al 2023 è stata Funzionario Archivistica di Stato in servizio presso l'Istituto Centrale per il Catalogo e la Documentazione dove ha ricoperto i seguenti incarichi: coordina-trice dell'Area Catalogazione dell'ICCD e Responsabile della gestione documentale e della conservazione digitale. Ha curato in particolare gli aspetti legati alla valorizzazione e alla condivisione dei dati del Catalogo generale dei beni culturali. È autrice di pubblicazioni e saggi sul tema della digitalizzazione e delle tecnologie del web semantico applicate al patrimonio culturale.

## Notes

[1] Si tratta di modelli concettuali, che esprimono la semantica sottesa alle attuali trenta normative catalografiche standard, per la strutturazione della conoscenza per i beni culturali. La rete di ontologie di ArCo ancora in via di sviluppo, si compongono di più moduli, tutti resi disponibili su GitHub con licenza aperta (<https://github.com/italia/daf-ontologie-vocabolari-controllati>). ArCo è allineata alle principali ontologie anche internazionali del dominio dei beni culturali (Europeana Data Model, Cidoc-CRM, Bibframe, CulturalON etc.) e fa parte ufficialmente – come ontologia per il patrimonio culturale – della rete di ontologie per la pubblica amministrazione rilasciate da AgID e definita OntoPiA.

[2] <https://www.catalogo.beniculturali.it/>

[3] Il sito viene gestito dall'ICCD per conto del MIC. Attraverso di esso i dati catalografici alimentano anche il portale governativo dei dati aperti e il portale europeo dei dati (<https://dati.gov.it>; <https://data.europa.eu/>)

[4] La procedura impiega RDFizer, un modulo software che converte automaticamente i dati basati sugli standard ICCD dal formato XML in RDF. Il software è disponibile gratuitamente sul canale GitHub di ICCD, secondo le disposizioni del CAD sul riuso del software per la pubblica amministrazione (artt. 68 e 69).

[5] Il gruppo che ha lavorato alla progettazione del nuovo sistema è composto da: Ada Gabucci, Iulia Fioravanti, M. Letizia Mancinelli, M. Letizia Melone, Riccardo Montalbano, Elena Musumeci, Cecilia Perrone, Chiara Veninata.

[6] La valutazione delle esigenze dell'utenza, attraverso interviste e un questionario sottoposto a un target di utenza con profili differenziati, ha consentito di perfezionare il processo di progettazione, formalizzato con l'elaborazione di User Story per la definizione dei requisiti del software e del product backlog.

[7] L'espressione "catalinking" è stata introdotta nel gennaio 2013 da Eric Miller al meeting dell'American Library Association, dalle parole "Cataloguing" e "Linked Data". L'impiego di strumenti di AI nello sviluppo del nuovo Sistema del Catalogo Nazionale si orienta principalmente a supportare il catalogatore nei processi di catalinking e di verifica del dato descrittivo.

[8] <https://digitallibrary.cultura.gov.it/il-piano/>

[9] <https://ipac.cultura.gov.it/>

[10] Troviamo qui anche i servizi che garantiscono la gestione e l'automazione dei workflow o l'interoperabilità tramite API Gateway e l'integrazione con altri sistemi del MIC e con i servizi dell'IPaC.

[11] Il processo prevede il ricorso a un harvesting OAI-PMH che espone file XML e a regole di trasformazione che rendono accessibili i LOD tramite uno SPARQL endpoint. L'harvesting viene usato per l'interoperabilità con CulturalItalia e come base per la trasformazione in linked open data e la successiva pubblicazione sul portale di consultazione del Catalogo.

[12] SHACL Shapes Constraint Language, è un linguaggio di specifica dei vincoli raccomandato dal consorzio W3C per validare dati in RDF in conformità con i requisiti specifici di dominio (<https://www.w3.org/TR/shacl/>).

# Geoportale Nazionale per l'Archeologia: la carta archeologica del territorio nazionale alla prova della sostenibilità

Valeria Boi<sup>1</sup>, Ada Gabucci<sup>2</sup>, Francesco Marucci<sup>3</sup>

<sup>1</sup>Istituto Centrale per l'Archeologia, <sup>2</sup>Cooperativa ALVEO di Bologna (responsabile Template QGIS), <sup>3</sup>Cooperativa ALVEO di Bologna (responsabile Geodatabase e WebGIS)

**Abstract.** The National Geoportal for Archeology (GNA) is a project started in 2017 by the Central Institute for Archeology of the Italian Ministry of Culture, as part of the activities carried out by the Institute regarding the census and digitalization of archaeological research data. The Geoportal constitutes the online collection and sharing point of data resulting from archaeological investigations carried out on the national territory. Its primary goal is the creation of a dynamic archaeological map, easily implementable over time, free to access and easy to consult, open to reuse and integration by all users. The data-entry according to the GNA standard is entrusted to the use of a pre-set GIS project (template) developed using the open-source software QGIS. The user stores information relating to the context of data collection, the type of intervention and the actual archaeological data, relating both to what is physically preserved and to any reconstructive hypotheses.

**Keywords.** Geoportale Nazionale per l'Archeologia (GNA), Open Data, Open Access, Cultural Heritage Protection, WebGIS, Open Source software

## Introduzione

L'ICA (Istituto Centrale per l'Archeologia) ha avviato la progettazione del Geoportale Nazionale per l'Archeologia (GNA) nel 2018, ponendosi come obiettivo la costruzione di una carta archeologica dinamica del territorio italiano. La prima fase del progetto, condotta sulla documentazione d'archivio di alcune Soprintendenze, è servita per la sperimentazione della struttura del database e delle modalità di inserimento dei dati, ma in pochi anni lo standard GNA è stato esteso a tutti i nuovi dati raccolti dagli Uffici del Ministero della Cultura - MiC (Calandra et al. 2021, Calandra 2023). Il portale è online da luglio 2023 (<https://gna.cultura.gov.it>), e oggi l'invio al GNA è parte integrante delle norme di consegna della documentazione di tutte le indagini archeologiche. Questa previsione, sancita da alcune Circolari del MiC (DG ABAP n. 9/2024 e DGABAP-DG-MU n. 22/2024, [https://gna.cultura.gov.it/wiki/index.php?title=Istruzioni\\_operative](https://gna.cultura.gov.it/wiki/index.php?title=Istruzioni_operative)), riguarda gli studi svolti nell'ambito della progettazione delle opere pubbliche (archeologia preventiva e "assistenze in corso d'opera"), ma anche scavi svolti per lavori privati o direttamente dal MiC, e integra una Circolare precedente (DG ABAP-DG MU 47/2022, <https://ica.cultura.gov.it/concessioni-di-ricerca/>) che disciplina le ricerche svolte dalle Università (Falcone 2023).

L'idea progettuale è nata dall'esigenza di mettere a disposizione di tutti uno strumento agile di consultazione di dati di base: interpretazione, descrizione e cronologia dei ritrovamenti, attendibilità del posizionamento, stato di conservazione; considerate le preminenti finalità di tutela, anche la “mancanza di ritrovamenti” è un dato rilevante: il sistema, dunque, archivia sempre esito e profondità delle indagini, anche se negative.

Nel panorama della tutela archeologica i progetti esistenti, anche in ambito ministeriale (fra tutti il Catalogo generale dei beni culturali, <https://catalogo.cultura.gov.it/>), sono incentrati su un esame molto approfondito dei rinvenimenti, affidato a specialisti e svolto in un momento successivo rispetto alle indagini sul campo; problemi principali cui il GNA ha tentato di dare una risposta sono stati, invece, la rapidità di aggiornamento e la sostenibilità: per questo motivo, la raccolta dei dati è affidata a chi conduce le attività sul campo, e i dati sono pubblicati online entro 30 giorni dall'invio al sistema. Tutti i dati sono distribuiti in formato aperto con licenza CC BY 4.0 direttamente dal portale, tramite download diretto o mediante servizi OGC.

Con la volontà di non disperdere il patrimonio di dati digitali raccolti negli ultimi 15 anni da progetti elaborati da singoli Uffici del Ministero, protocolli di interoperabilità permettono ai sistemi preesistenti di continuare a operare, conferendo al GNA i dati minimi, senza modificare i flussi di lavoro già testati sul territorio; è questo il caso delle “banche dati federate”, ognuna delle quali scambia dati con il GNA secondo strategie studiate caso per caso ([https://gna.cultura.gov.it/wiki/index.php?title=Istruzioni\\_operative](https://gna.cultura.gov.it/wiki/index.php?title=Istruzioni_operative), par. 5).

Nell'ottica di creare un vero e proprio “hub di dominio”, inoltre, il GNA accoglie in una sezione separata anche progetti nati in tempi e con finalità differenti, per i quali non è prevista una mappatura dettagliata, ma solo la condivisione delle informazioni (<https://gna.cultura.gov.it/mappa.html?sezione=altrebanchedati>).

## 1. Inserimento e invio dei dati

L'immissione dei dati avviene attraverso un Template QGIS distribuito gratuitamente online, che costituisce uno strumento costantemente aggiornato e che garantisce l'uniformità della raccolta dei dati. L'utilizzo di un progetto in locale, inoltre, lascia ai singoli la libertà di utilizzare tutta la documentazione necessaria, eventualmente anche in formati non compatibili con GNA, e di utilizzare tutta la vasta gamma di strumenti messi a disposizione dal software. La prima versione del Template è stata distribuita in forma sperimentale nel 2019 e, dopo alcune release, nel 2022 si è arrivati allo standard per la raccolta dei dati degli studi preliminari nell'ambito dell'archeologia preventiva (Gabucci 2023).

Con raffinamenti successivi, a cui hanno contribuito tutti i funzionari e professionisti che hanno lavorato durante la sperimentazione e nelle prime fasi della messa in produzione, si è giunti a una versione abbastanza matura del progetto (1.4.2, <https://gna.cultura.gov.it/download.html>).

Il rilascio più recente sistematizza ulteriormente varie informazioni, con l'obiettivo di una ulteriore normalizzazione, anche dei nomi di compilatori/responsabili/funzionari Mic competenti: operazione, quest'ultima, funzionale anche a garantire l'invio automatico dei dati tramite plugin ai funzionari incaricati della tutela dei diversi territori.

Modifiche sostanziali sono state apportate anche alla modalità di registrazione della cronologia. Fin dall'inizio della progettazione, si è scelto di privilegiare la definizione in date assolute espresse da numeri reali, per consentire ricerche e confronti, anche in ambito internazionale. Con l'intento di semplificare il lavoro dei compilatori, però, i range cronologici sono stati inizialmente legati a definizioni testuali di macro-fasi cronologiche; questa scelta si è scontrata con la estrema frammentazione dell'area italiana in età antica, che determina in letteratura il proliferare di "etichette" in cui la connotazione cronologica è strettamente congiunta a quella culturale ("facies di Monte Claro", "età di Pirro", Facies di Golasecca"), in un fiorire di definizioni praticamente impossibili da normalizzare e da inserire in un vocabolario chiuso e standardizzato, ma che gli studiosi sono sempre stati restii a far "assorbire" dalle macro-fasi codificate dal GNA.

Per queste ragioni, siamo giunti oggi a rendere obbligatoria solo l'indicazione degli estremi cronologici assoluti, chiedendo al compilatore di indicare in un ulteriore campo testuale la definizione di una eventuale etichetta, unita a un riferimento bibliografico/fonte dell'attribuzione del range cronologico. Il risultato è quello di restituire in banca dati la complessità del contesto, pur non rinunciando alla registrazione del dato cronologico assoluto; in secondo luogo, la costruzione di un thesaurus di definizioni con valenza locale può essere preziosa per stimolare il confronto fra gli studiosi, ma anche, a livello internazionale, per poter contribuire a progetti come PeriodO (<https://perio.do/en/>), in cui oggi tali definizioni non sono attestate.

## 2. Database e plugin qgis

Il cuore del sistema GNA è costituito dal Database, nel quale i dati e relazioni sono strutturati per accogliere, analizzare e rappresentare tutte le informazioni da gestire.

Una parte importante dei dati è formata da informazioni geografiche georeferenziate espresse in punti, linee e poligoni. Il geo-database quindi è strutturato in modo da contenere anche le relazioni spaziali tra i dati, espresse in funzioni geometriche, che permettono di effettuare tutte le operazioni GIS necessarie ad assicurare la coerenza tra le informazioni. I punti di ingresso dei dati nel database sono molteplici, in generale dipendenti dai diversi ruoli degli attori del GNA: professionisti, funzionari, amministratori del sistema, ma anche cittadini.

Lo strumento principale che permette l'immissione dei dati è costituito dal già citato Template, per il quale è stata sviluppata una interfaccia grafica (plugin) che permette all'utente di inviare i dati al GNA. La consegna dei dati viene sempre assegnata ad uno o più funzionari. Una volta ricevuto un pacchetto di dati, il funzionario ha a disposizione un Template e un plugin dedicati che gli consentono di gestire i dati a lui assegnati: è stato previsto quindi un percorso mediante il quale, attraverso la modifica dello stato e la scadenza dei termini di attesa, il dato consegnato (se necessario modificato) viene pubblicato in GNA e quindi reso disponibile per tutti.

Altri punti di ingresso "privilegiati" sono utilizzati dalle procedure di inserimento e aggiornamento dei dati provenienti dalle banche dati federate, ovviamente dopo essere stati normalizzati e strutturati nello standard GNA.

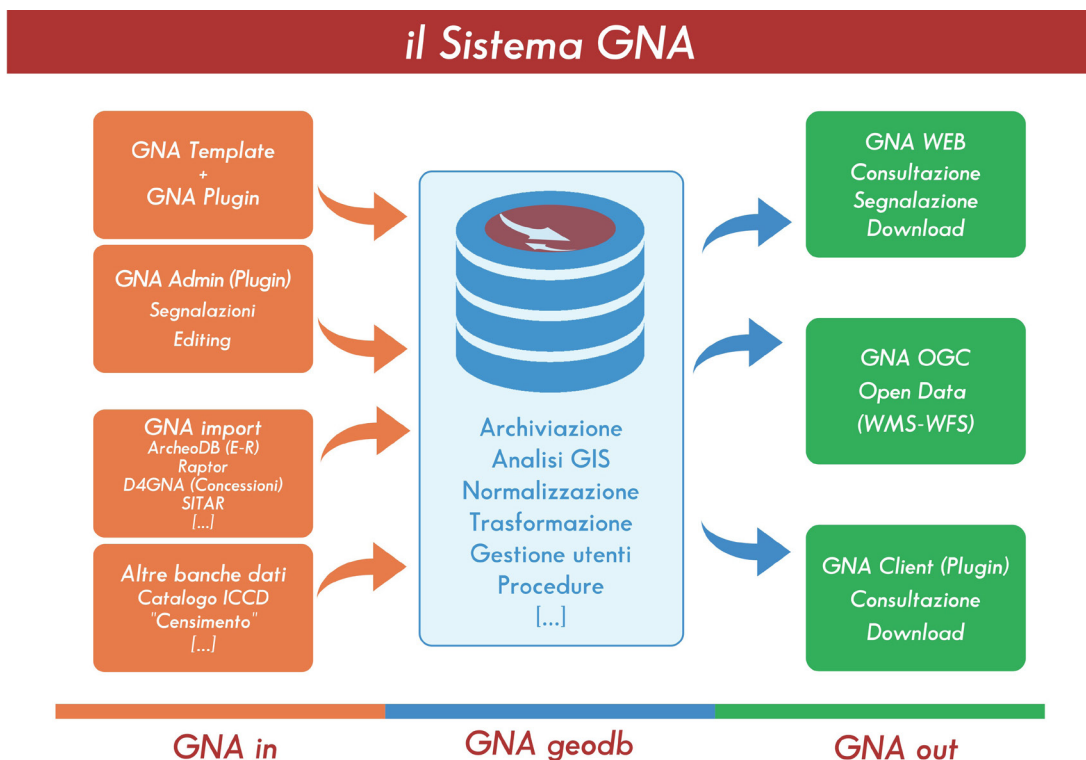
Il geo-database quindi, in tutti i casi precedenti, ha la funzione di archiviazione, di analisi geometrica (per assicurare la coerenza nelle relazioni geografiche, tipicamente l'appartenenza alle unità amministrative), di armonizzazione e normalizzazione per la trasformazione nello standard; infine, contiene anche tutte le procedure automatiche per la gestione delle informazioni.

Dal punto di vista della rappresentazione dei dati, il Sistema GNA prevede anche qui diversi strumenti:

- il portale WEB, inteso come riferimento per la consultazione dei dati nella mappa interattiva (WebGIS), la possibilità di effettuare delle interrogazioni guidate sui dati;
- i servizi WEB che permettono lo scaricamento del dato trattato secondo gli standard OGC sia come scaricamento del risultato delle query in un pacchetto vettoriale;
- il Plugin per il Template che permette di integrare nel proprio modulo di progetto i dati presenti nel GNA al momento dell'integrazione.

In sintesi, le tre parti del sistema GNA possono essere schematizzate in GNA in, GNA geodb e GNA out, come rappresentato nella figura 1.

Fig. 1  
Lo schema del Sistema GNA (F. Marucci)



### 3. Conclusioni

L'adozione del GNA come standard nazionale rappresenta un significativo passo avanti verso la costruzione di una carta archeologica realmente dinamica e aggiornata in tempo reale. La ricerca della sostenibilità del progetto ha orientato verso una soluzione "collaborativa" che mette in primo piano il ruolo del singolo utente nel registrare, aggiornare e

condividere i dati. Questa scelta porta in primo piano temi di grande rilevanza, come la “proprietà” e la “paternità” intellettuale dei dati pubblicati, i diversi ruoli svolti da professionisti, studiosi, funzionari ministeriali. la gestione di modifiche/aggiornamenti da parte di persone diverse dall'autore della prima scheda.

## Riferimenti bibliografici

Acconcia, V. (2023), La pubblicazione dei dati nel Geoportale Nazionale per l'Archeologia. doi:10.60974/GNA\_03

Boi, V. (2023), Il Geoportale Nazionale per l'Archeologia (GNA). Standardizzazione e apertura dei dati. doi:10.60974/GNA\_02

Calandra, E. (2023), Il Geoportale Nazionale per l'Archeologia (GNA). Un'introduzione. doi:10.60974/GNA\_01

Falcone, A. (2023), D4GNA. Dematerializzazione e condivisione in rete dei dati delle indagini archeologiche svolte in regime di concessione. Promozione e condivisione dei dati prodotti dalle missioni archeologiche italiane all'estero. doi:10.60974/GNA\_04

Gabucci, A. (2023), Un Template QGIS al servizio del Geoportale Nazionale per l'Archeologia (GNA). doi:10.60974/GNA\_05

Calandra, E., Boi, V., Falcone, A., Acconcia, V., Di Giorgio, S., Massara, F. e Ronzino, P., (2021), Policy and Practice for Digital Archaeological Archiving in Italy, in Internet Archaeology 58, 2021. Articolo: <https://doi.org/10.11141/ia.58.27>

## Autori

Valeria Boi [valeria.boi@cultura.gov.it](mailto:valeria.boi@cultura.gov.it)

Dottore di ricerca in metodologia della ricerca archeologica; funzionario archeologo presso l'ICA, è coordinatore del GNA; si occupa in particolare di standard descrittivi, flussi documentali e interscambio dati, e delle modalità di pubblicazione e riuso dei dati; segue le relazioni con gli uffici periferici per gli interventi diretti dal Mic e le iniziative di formazione. Fa parte del Comitato Scientifico del Convegno Internazionale Archeofoss e della rivista Open Access Archeologica Data.



Ada Gabucci [ada.gabucci@gmail.com](mailto:ada.gabucci@gmail.com)

Dottore di ricerca in archeologia, specialista per l'età romana. Si occupa di razionalizzazione dei dati archeologici e della loro strutturazione e più in generale della catalogazione dei beni culturali e delle normative ministeriali, con esperienza nel campo della proposizione di liste e di vocabolari e di progettazione e realizzazione di sistemi informativi territoriali anche complessi.

In qualità di archeologa professionista ha esperienza trentennale di consulenza al Mic e ad altri Enti pubblici.

Francesco Marucci [francesco.marucci@gmail.com](mailto:francesco.marucci@gmail.com)

Sviluppatore esperto in Sistemi Informativi Geografici, si occupa soprattutto di progettazione di banche dati geografiche e layout cartografici, Sviluppo di applicativi software geografici (desktop, WEB e Mobile), implementazione di geoDatabase in ambiente OpenSource, ed elaborazione, trasformazione e analisi di dati geografici.

# La Piattaforma di big data in Istat nei processi statistici e di sperimentazione

Francesco Altarocca, Marco Polizzi

Istat

**Abstract.** Negli ultimi anni, in conformità con i Piani Triennali per l'informatica interni, l'Istat ha applicato una politica per garantire l'evoluzione continua, la razionalizzazione e il consolidamento delle piattaforme IT, nell'ottica di ottimizzare il supporto tecnologico alla ricerca e ai processi di produzione statistica. In particolare, l'Istat ha adottato soluzioni IT on premise, per la cui gestione ed evoluzione è stato necessario coinvolgere figure specializzate (sistemisti, architetti, analisti), in grado di interfacciarsi con esperti di dominio e sviluppatori in soluzioni big data. Il presente contributo descrive l'esperienza Istat nell'acquisizione di una piattaforma big data, sulla quale sono stati implementati importanti processi di produzione statistica. La piattaforma ha consentito di svolgere alcune sperimentazioni su algoritmi computazionali particolarmente complessi, che hanno permesso ai ricercatori di sviluppare metodologie statistiche prima difficilmente implementabili

**Keywords.** Big data, tecnologie abilitanti, data lake, basi di dati a grafo

## Introduzione

Nell'ambito delle attività di ricerca nelle quali è impegnato, l'Istat cura anche lo sviluppo di tecnologie innovative abilitanti alle nuove fonti di dati.

La disponibilità di grandi quantità di dati o big data (in seguito BD), con particolare riferimento a quelli della Grande Distribuzione Organizzata (in seguito GDO), è cresciuta considerevolmente e ha evidenziato la necessità di adeguare i sistemi IT per far fronte ai nuovi carichi elaborativi, abilitando contestualmente i ricercatori a nuovi scenari di indagine.

Il presente contributo descrive l'esperienza Istat nell'acquisizione di una Piattaforma Big Data (in seguito BDP), sulla quale sono stati implementati importanti processi di produzione statistica. La piattaforma ha inoltre reso possibile la sperimentazione di algoritmi complessi a livello computazionale, consentendo di sviluppare metodologie statistiche prima difficilmente implementabili.

## 1. I processi di produzione statistica con la BDP

### 1.1 Gli scanner data

Annualmente l'Istat rivede l'elenco dei prodotti che compongono il paniere di riferimento della rilevazione dell'Indice dei Prezzi al Consumo (in seguito IPC), anche per la misura dell'inflazione. Dal 2018 l'Istat utilizza i prezzi registrati alle casse della GDO (AA.VV. 2018) mediante scannerizzazione dei codici a barre o scanner data (in seguito SD).

## 1.2 L'indice dei prezzi al consumo

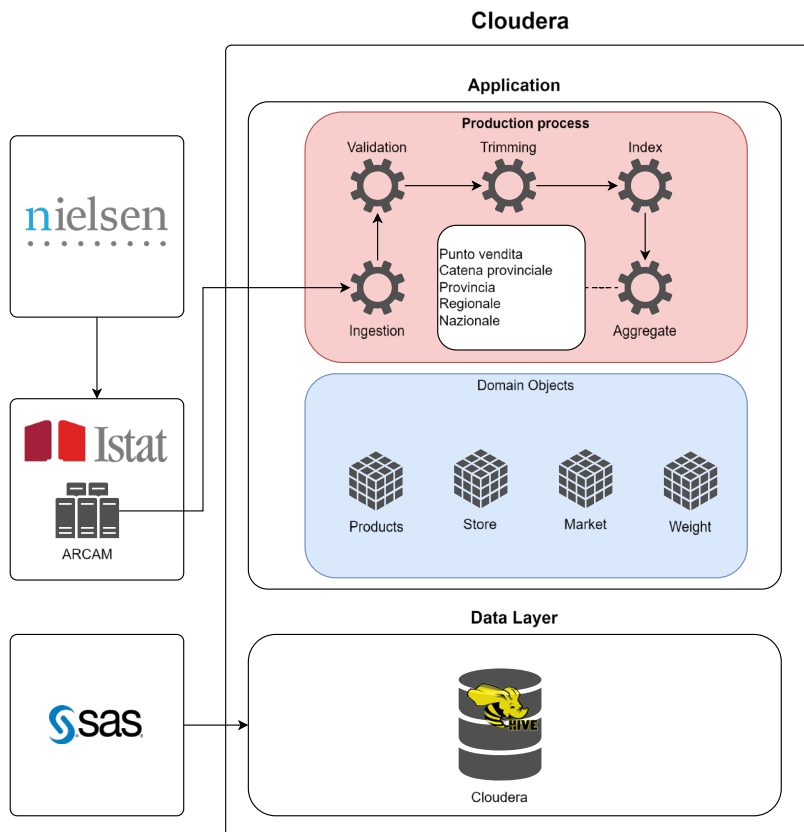
Negli ultimi anni, il calcolo della stima dell'IPC è stato contraddistinto da un sostanziale aumento delle referenze prese in esame e del numero dei punti vendita da inserire nel campione. I database relazionali impiegati in precedenza hanno manifestato alcune criticità, fra le quali:

- la difficoltà di accogliere e scalare su una quantità di dati sempre crescente;
- la complessità nell'immagazzinare ed elaborare più annualità della stessa rilevazione;
- la difficoltà a raggiungere performance adeguate a svolgere in modo efficiente attività di controllo sulle forniture.

La BDP utilizzata in Istat ha consentito di risolvere tali problematiche, favorendo un'implementazione ottimale del processo di produzione statistica dell'IPC. Settimanalmente l'Istat acquisisce i dati di fatturato, quantità e tipologia dei prodotti dalla GDO per tutte le province italiane. Il campione dei punti vendita è rappresentativo dell'universo delle tipologie distributive della GDO e comprende circa 4.000 punti vendita distribuiti su tutto il territorio nazionale. Nel solo 2023 sono stati acquisiti circa 1 miliardo di dati provenienti dalla GDO.

Per ciascun prodotto e per ciascun punto vendita viene creato e gestito un micro-indice, per un totale di 35 milioni di micro-indici mensili e 421 milioni annuali. Sulla base dei micro-indici sono determinati, mensilmente, quasi 7 milioni di aggregazioni per livello territoriale (provinciale, regionale e nazionale), per raggruppamento di mercato e per tipologia di GDO, per un totale annuo di 80 milioni di aggregati (dati relativi al 2023).

Fig. 1  
Architettura del sistema per l'elaborazione degli scanner data



Nella Figura 1 è riportato lo schema del flusso di elaborazione dei dati necessario al processo SD per la produzione dell'IPC.

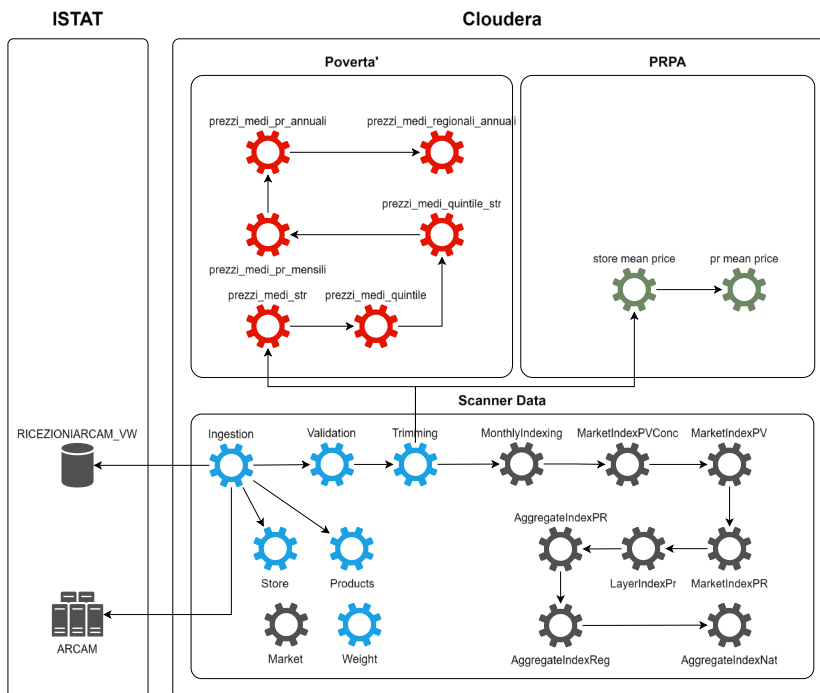
### 1.3 La stima della povertà e la parità regionale del potere d'acquisto

Le elaborazioni necessarie alle indagini sulla stima della povertà e della Parità Regionale del Potere d'Acquisto (in seguito PRPA), per la parte SD, sono realizzate utilizzando i dati elementari già acquisiti per l'indagine sugli IPC.

Fra i livelli considerati per la stima della povertà, oltre a quelli già utilizzati per l'IPC, viene incluso il formato del prodotto e sono presi in esame solo i formati di confezionamento più rappresentativi.

La BDP è utilizzata anche per il progetto sperimentale sulla PRPA, che fornisce una misura delle differenze nel livello medio dei prezzi di un paniere di prodotti fra una regione e l'altra e consente, quindi, di rilevare le diseguglianze e le condizioni di vita delle famiglie nei diversi territori dovute alle differenze nel potere d'acquisto che le caratterizza (AA.VV. 2023).

Fig. 2  
Package per il  
trattamento dei  
dati del data-  
lake degli scanner  
data



La Figura 2 schematizza i moduli software utilizzati in comune per le tre indagini appena descritte.

### 1.4 L'infrastruttura tecnologica big data in Istat

L'Istat ha adottato una soluzione di BD innovativa, implementando un'architettura Cloudera su un ambiente virtuale, basato sulla tecnologia iper-convergente Nutanix. Tale tec-

nologia consente di sfruttare alcuni vantaggi caratteristici degli ambienti virtuali, quali la flessibilità, la scalabilità, la facilità di gestione e la possibilità di ottenere elevate prestazioni. La BDP risiede su cluster composto da 21 macchine virtuali, 40 CPU, 64 TB di spazio disco e 2.330 GB di RAM.

## **2. Le sperimentazioni per lo sviluppo di metodologie statistiche**

### **2.1 Il record linkage probabilistico**

Oltre ai processi di produzione statistica appena descritti, la BDP ha reso possibile lo sviluppo di alcune interessanti sperimentazioni.

La sperimentazione sul record linkage probabilistico ha avuto lo scopo di effettuare alcuni test di performance e di fattibilità rispetto all'attività di record linkage con il progetto Splink, sfruttando le potenzialità della BDP. Splink è un progetto libero che, utilizzando le API Spark Python, implementa il modello Fellegi-Sunter per il record linkage (Linacre, R., et al. 2022). Nel corso dei test sono stati elaborati dataset di varie dimensioni, analizzati i tempi di esecuzione e individuata la configurazione migliore in base ai parametri utilizzati.

L'analisi dei risultati ottenuti si è rivelata utile per lo studio di fattibilità di altri progetti per i quali si prevede l'utilizzo della stessa tecnologia per elaborare una quantità di dati finora difficilmente gestibile.

### **2.2 Le basi di dati a grafo**

Un'altra sperimentazione è stata condotta sul modello del registro base degli individui, che costituisce il riferimento unico per tutte le statistiche ufficiali riferite alla popolazione abitualmente dimorante.

La sperimentazione ha avuto lo scopo di indagare le potenzialità dei database a grafo nel collegare entità e caratteristiche degli individui in layer differenti e di valutare potenziali punti di forza rispetto ai database relazionali. Per tale sperimentazione è stata utilizzata la libreria GraphX (Gonzalez, J. E., et al. 2014) per il calcolo parallelo sulla BDP in modalità trasparente.

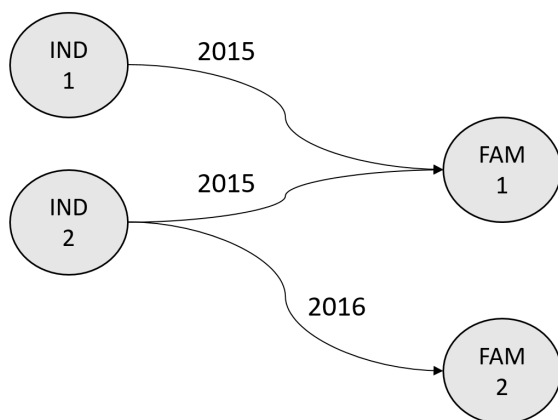
Il dataset di input era costituito da circa 3 milioni di record, dove ciascun record identifica un individuo e la relativa appartenenza a una famiglia durante un quinquennio.

Sono stati quindi prodotti vertici per rappresentare individui e famiglie, mentre gli archi identificano l'appartenenza di un individuo ad una famiglia per ciascun anno (Figura 3).

L'esito della sperimentazione ha evidenziato la facilità e l'efficacia con cui è possibile estrarre la struttura, le relazioni, le particolarità del tessuto delle famiglie e gli spostamenti di individui da una famiglia ad un'altra.

Un ulteriore promettente sviluppo della soluzione prevede l'integrazione di ontologie per aggiungere una componente di reasoning al sistema.

Fig. 3  
Esempio di modellazione di individui e di famiglie per la sperimentazione



### 3. Conclusioni

Partendo dai dati elementari del data-lake IPC, la BDP è stata impiegata per il parziale riutilizzo di pacchetti, procedure e flussi di lavoro utili allo sviluppo di altre indagini (cfr. Figura 2). La capacità della BDP di definire strutture dati indipendentemente dal formato, dalle partizioni e dalle posizioni dei dati sottostanti ha agevolato notevolmente i lavori per integrare altri processi produttivi e di ricerca.

Abilitare i ricercatori e gli esperti di dominio all'uso di tali tecnologie per ottenere vantaggi, sia rispetto alla velocità dei risultati, sia sulla capacità di gestione di grandi quantità di dati, sta diventando sempre più un aspetto fondamentale nella produzione di ricerca e nell'innovazione dei processi.

### Riferimenti bibliografici

AA.VV. (2018) GLI INDICI DEI PREZZI AL CONSUMO. Aggiornamenti del paniere, della struttura di ponderazione e dell'indagine – Nota informativa ISTAT, [https://www.istat.it/it/files//2018/02/Nota-informativa\\_-paniere2018\\_fp-1.pdf](https://www.istat.it/it/files//2018/02/Nota-informativa_-paniere2018_fp-1.pdf)

AA.VV. (2023) Indici spaziali dei prezzi al consumo - Anno 2021 – ISTAT, Testo integrale e Nota metodologica <https://www.istat.it/it/archivio/287297>

Gonzalez, J. E., Xin, R. S., Dave, A., Crankshaw, D., Franklin, M. J., & Stoica, I. (2014). {GraphX}: Graph processing in a distributed dataflow framework. In 11th USENIX symposium on operating systems design and implementation (OSDI 14) (pp. 599-613)

Linacre, R., Lindsay, S., Manassis, T., Slade, Z., & Hepworth, T. (2022). ApplyiSplink: Free software for probabilistic record linkage at scale. *International Journal of Population Data Science*, 7(3), 1794. <https://doi.org/10.23889/ijpds.v7i3.1794>

## Autori



**Francesco Altarocca** [fraltaro@istat.it](mailto:fraltaro@istat.it)

Ricercatore presso l'Istat (Dir. Informata DCIT) da oltre 20 anni. È responsabile dell'iniziativa "Produzione sistemi IT per Big Data" ed è coinvolto in linee di attività, progetti strategici e comitati in ambito big data, nuove fonti, AI, strumenti innovativi e Trusted Smart Statistics. Coordina task force e gruppi di lavoro, ha effettuato docenze nell'ambito del corso "La PA nell'epoca dei big data" e collabora con altri autori per articoli e pubblicazioni.

**Marco Polizzi** [polizzi@istat.it](mailto:polizzi@istat.it)

Tecnologo presso la Direzione centrale per le tecnologie informatiche – Istat

Ha svolto attività di ricerca su remote sensing presso l'Università "La Sapienza" di Roma e l'Enea.

Come Istat è stato responsabile dell'iniziativa IT "Architetture dati, Big Data, LOD e BI", ora dell'iniziativa IT "Piattaforme per gestione e analisi dati", per il supporto informatico alle indagini statistiche con strumenti RDBMS e BI. È stato referente Istat per il progetto PNRR Catalogo Nazionale Dati.



# On Data Throughput in Real-Time Structural Health Monitoring: a Next-Generation EU funded Project use case

Paolo Bellagente

University of Brescia – Department of Information Engineering

**Abstract.** Breve Structural Health Monitoring (SHM) has started to be used to gain a deeper knowledge of infrastructure conditions and the evolution of transport infrastructures, as their development and maintenance are crucial to ensure the wealth and safety of human activities. The goal of this paper is to quantify the order of magnitude of the amount of data produced by an SHM project using a real test case in northern Italy as a reference.

**Keywords.** Digital Twin, Structural Health Monitoring, Research data, Network throughput

## Introduction

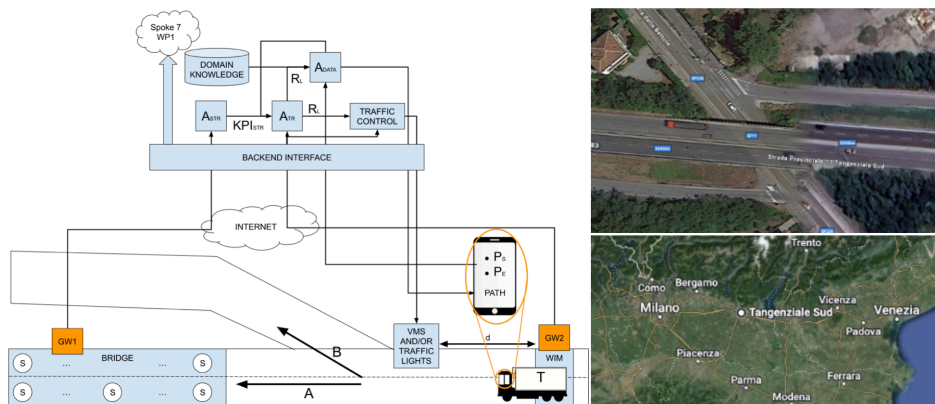
Mobility systems constitute a fundamental element for the efficiency and sustainability of social, economic, and territorial frameworks. Large-scale infrastructural networks are essential for the seamless flow of long-range traffic. The intricacy of mobility systems is likely to escalate with the proliferation of cooperative, connected, and automated mobility, the dissemination of innovative mobility models, and the advent of novel energy paradigms. Despite their strategic significance, physical infrastructures often have antiquated features, thereby raising concerns regarding their efficiency and safety in numerous countries. Bridges, viaducts, and tunnels concurrently represent vulnerable assets within the mobility network and usually lack convenient alternative resources or routes. Localized failures in these components quickly propagate, adversely affecting network efficiency and potentially causing severe consequences. To limit such impacts, Structural Health Monitoring (SHM) has started to be used to gain deeper knowledge of infrastructure conditions and evolution (Wang G., Ke J., 2024). Particularly for bridges, forecasting the severity of events where design traffic load limits have been exceeded is fundamental for road safety (Ventura R., et al, 2024). Fusing information about the kind of load that traffic imposes on a particular bridge, and how the structure reacts to such load, will make it possible to develop a forecasting model of the bridge. This model can be used to control traffic, making heavy vehicles avoid the bridge and choosing alternative routes to reach their destinations. The feasibility study of such a control system and its impact on the lifecycle extension of a bridge is the University of Brescia's in-progress contribution to the National Centre for Sustainable Mobility, funded by the Italian National Recovery and Resilience Plan (PNRR) with Next Generation Europe grants. One of the problems faced

in this project regards collecting data from instrumentation applied to the bridge and analysing it, looking for useful insight to develop the digital twin of the infrastructure. Currently, there is a clear absence of established standards for evaluating the dependability of digital twins (Tran M. Q., et al, 2024) . The proliferation of sensors on structures, while favourable to enhanced monitoring, incurs exorbitant costs. Hence, there exists a pressing need for further optimization in the scopes of data generation and predictive modelling. However, trying to develop such models (Zar A., et al.,2024)(Rehman S. U., et al, 2024) (Valkonen A., Glisic B., 2023) results in the production of a massive amount of data (Fleischhacker A., et al, 2020) as it is impossible to filter data in advance when what is demanded is unknown (Pasetti M., et al.,2021). Thus, the goal of this paper is to estimate of the communication network load generated by an SHM research project, using the use case of the University of Brescia, as they will be relevant in the Italian National landscape as demonstrated by the number of funds recently allocated (EUR-Lex, 2023)(Council of the European Union, 2021).

### 1. The Use Case

The system proposed by the University of Brescia will implement a method for the real-time assessment and management of the risk induced by heavy traffic loads passing over bridges. This method is based on the use of Weight-In-Motion (WIM) technologies for measuring the load in transit and sensors for monitoring the structural response. Fig. 1 shows a functional scheme of the proposed conceptual system and the localisation of the considered bridge in northern Italy.

Fig. 1  
Functional scheme of the proposed system and its localisation in northern Italy



Specifically, a data collection system for the characteristics of passing vehicles (WIM station) and a data collection system for structural response will be implemented on the selected pilot bridge on the SP BS EX SS 11 Tangenziale Sud freeway, in agreement with the local Road Authority. Data collected by sensors, deployed in proximity of the bridge, will be used to simulate the scenario of Fig. 1. In this scenario, the driver of a heavy vehicle transmits its start location and its destination to the Analysis Backend where a routing algorithm (Adata in Fig. 1) returns an appropriate route optimising the requested

truck mission with respect to the traffic congestion and the road network and the risk of reaching specific Limit States associated with the structural response of the bridge. For example, Ultimate Limit States (ULS) related to reaching the design load of the structure or Serviceability Limit States (SLS) related to reaching deformability or crack formation limits (Ventura R., et al, 2024).

The system keeps structural KPIs and ULS/SLS risk factor up-to-date in real-time. For each vehicle transiting on the WIM, its weight, speed and classification are immediately sent to the Analysis Backend and elaborated by the Transportation Analysis block (ATR in Fig. 1). Knowing the vehicle speed (which doesn't vary much for heavy vehicles on a freeway) and synchronising the WIM and the SHM systems, the latter will record the structural reaction of the bridge under the solicitation posed by the heavy vehicle. These data are sent to the Analysis Backend to be elaborated by the digital model of the bridge (ASTR in Fig. 1). The last goal is to simulate an experimental system to protect the bridge in case of a high-risk level, limiting the structural consequent damages and lowering the time and costs of restoration. The idea is to install the WIM system in a suitable position before a freeway exit and install a Varying Message Signal (VMS) to communicate to the driver of the dangerous vehicle that it must exit as the passage on the bridge is forbidden.

## 2. Data Sources

The measuring instrumentation will be provided by iWIM srl. The SHM system is composed of optical inclinometers and accelerometers suitably installed on the bridge. For each vehicle measured by the WIM, the measure will be encoded in a structured JSON and sent to a specific URL via an HTTPS POST request. The JSON contains vehicle weight, speed, counting and classification, Number of axles, single axle shaft weight, first/last axis distance, axle width, twin wheel axle presence, 3 images encoded as a base64 string: a colour 1920x1080 image of the context camera, a black and white 1920x1080 image of the number plate reading camera and a detail 198x74 image of the number plate. The total size of each JSON register is approximately 300 KB on average. Given that the average traffic passing on the freeway is approximately 18,600 vehicles/day, the average data flow produced by the WIM system is 5.6 GB.

The SHM system provides the real-time transmission of 19 engineering quantities encoded in double-precision floating-point format (8 bytes) and a time reference. The maximum sampling frequency of the system is 500 Hz producing a binary flow of 7.0 GB/day. It is possible to optimize the SHM data sampling using the WIM data, reducing the sampling frequency when a load is not expected. A statistical analysis of the traffic recorded for the considered freeway suggests the daily presence of approximately 150 "notable" (over 50t) transits which will trigger a 500Hz acquisition for 1 Minute. Under non-critical load conditions, the data collected by the sensors will be sent at a reduced frequency of 5 Hz resulting in a reduced data flow of 1 GB/day. Unfortunately, such optimisation is not suitable for a scientific study as the real behaviour of the bridge is unknown and the 1-minute window could truncate interesting observations.

### 3. Conclusions

Development and maintenance of mobility systems are crucial to ensure the wealth and safety of human activities. For these reasons, scientific research has recently multiplied its effort in studying and testing SHM solutions to ensure an effective and secure functioning, asset management, and the provision of a secure and high-quality travel experience on the physical mobility infrastructure. To evaluate the amount of data produced by SHM systems in scientific research, the solution proposed by the University of Brescia has been described. The system composed of a WIM instrument and a SHM installed on a bridge of the south freeway in Brescia, Italy, is able to produce data for 12.6GB/day. As general models able to describe any bridge are not yet available, and the average life of transport infrastructure is tens of years, the number of SHM research projects is destined to rise and their duration overlaps. The amount of raw data generated by instruments should be routed to research centres and universities for the needed elaboration stressing the research network. In Italy, there are at least more than 21000 road bridges among which the one described as a use case is a small example. If no optimisation is found in the early stages of experiments the total amount of data collected will be, at least, 264.6TB/day only for bridges with few monitored parameters. Particular attention should be paid in the future to foresee possible orchestration in real-time raw data collection studies or to focus on modelling to improve the edge-computing paradigm in SHM applications.

### Acknowledgements

This work is part of the research activity developed by the authors within the framework of the “PNRR”: SPOKE 7 “CCAM, Connected Networks and Smart In-frastructure” - WP4

### Riferimenti bibliografici

Council of the European Union (2021), “ANNEX to the COUNCIL IMPLEMENTING DECISION amending Implementing Decision of 13 July 2021 on the approval of the assessment of the recovery and resilience plan for Italy.”

EUR-Lex (2023) 442 - EN - EUR-Lex.” Accessed: Mar. 12, 2024. [Online]. Available: [https://eur-lex.europa.eu/procedure/EN/2023\\_442](https://eur-lex.europa.eu/procedure/EN/2023_442)

Fleischhacker A., et al (2020) “Bayesian Survival Analysis for US Concrete Highway Bridge Decks,” *Journal of Infrastructure Systems*, vol. 26, no. 1, p. 04020001.

IWIM S.R.L. (2024), “iWIM Brochure,” Technical Data Sheet, [Online]. Available: <https://www.iwim.it/>

IWIM S.R.L. 2 (2024), “Bridge SHM Brochure,” Technical Data Sheet. [Online]. Available: <https://www.iwim.it/>

Nobile M., Samorì C. (2024), “Ainop e Linee Guida: Il Ruolo dei Dati per la Sicurezza

di Ponti E Viadotti,” *Ingenio*, <https://www.ingenio-web.it/articoli/ainop-e-linee-guida-il-ruolo-dei-dati-per-la-sicurezza-di-ponti-e-viadotti/> (accessed Mar. 13, 2024).

Pasetti M., et al. (2021), “Impact of the Measurement Time Resolution on Energy Key Performance Indicators for Distributed Energy Resources: An Experimental Analysis,” *AM-PS - 11th IEEE International Workshop on Applied Measurements for Power Systems, Proceedings*.

Rehman S. U., et al (2024) “Advancing structural health monitoring: A vibration-based IoT approach for remote real-time systems,” *Sensors and Actuators*, vol. 365, 2024.

Tran M. Q., et al (2024) “Opportunities and Challenges of Digital Twins in Structural Health Monitoring,” *Lecture Notes in Civil Engineering, LNCE* ,vol. 344, pp. 673–681.

Valkonen A., Glisic B. (2023) “Structural Health Monitoring-Based Bridge Lifecycle Extension: Survival Analysis and Monte Carlo-Based Quantification of Value of Information,” *Infrastructures*, vol. 8, no. 11, Art. no. 11.

Ventura R., et al (2024), “Prediction of the severity of exceeding design traffic loads on highway bridges”, *Heliyon*, vol. 10, no. 1

Wang G., Ke J. (2024), “Literature Review on the Structural Health Monitoring (SHM) of Sustainable Civil Infrastructure: An Analysis of Influencing Factors in the Implementation”, *Buildings*, vol.14, no.2

Zar A., et al. (2024), “Towards vibration-based damage detection of civil engineering structures: overview, challenges, and future prospects,” *International Journal of Mechanics and Materials in Design*.

## **Autori**



**Paolo Bellagente** [paolo.bellagente@unibs.it](mailto:paolo.bellagente@unibs.it)

Is a researcher in the Embedded Electronics and Smart Sensors group of the University of Brescia. Department of Information Engineering since 2018. He is an expert in IoT architectures and data collection technologies for the smart city. He is a Member of the IEEE I&M Society and the IEEE IoT Technical Community. He is involved in the PNRR National Sustainable Mobility Center, spoke 7 (Cooperative, Connected and Automated Mobility, Connected Networks and Smart Infrastructure).

# L'utilizzo delle infrastrutture digitali per la riqualificazione del patrimonio edilizio italiano: un caso studio

Giacomo Bizzarri, Enrica Boldrin, Laura Ferrari

Università degli Studi di Ferrara

**Abstract.** Lo scopo di questo paper è quello di presentare i benefici tecnico-economici che i servizi digitali possono apportare nell'ambito delle riqualificazioni energetiche, in particolare con riferimento all'utilizzo dei Digital Twin. Il modello digitale è stato utilizzato per stimare l'efficacia degli interventi di riqualificazione per raggiungere una riduzione dei consumi di energia primaria del 90%. Il modello digitale è stato successivamente utilizzato per verificare l'effettiva efficacia di tali interventi attraverso il monitoraggio dei dati di consumo per un periodo di tempo di tre anni.

**Keywords.** Riqualificazione, energetica, modello digitale, edifici tutelati

## Introduzione

P2Endure è un progetto di ricerca finanziato dall'Unione Europea nell'ambito del programma Horizon 2020. Il progetto, terminato nel 2020, mirava a sviluppare e implementare soluzioni innovative per la riqualificazione energetica degli edifici, con un focus particolare su tecnologie plug-and-play.

P2Endure ha rappresentato un passo importante verso la modernizzazione del patrimonio edilizio europeo, rendendolo più efficiente, sostenibile e confortevole per gli occupanti. Grazie alle sue soluzioni innovative e al suo approccio integrato, il progetto ha lo scopo di rivoluzionare il modo in cui vengono effettuate le riqualificazioni energetiche, rendendole più accessibili e meno invasive.

Uno degli aspetti più significativi del progetto è l'utilizzo di strumenti digitali avanzati, tra cui il Building Information Modeling (BIM), per mappare gli edifici esistenti e pianificare interventi di riqualificazione personalizzati ed efficienti. L'uso di BIM e altre tecnologie digitali per la pianificazione e l'implementazione delle riqualificazioni rappresenta un significativo passo avanti rispetto ai metodi convenzionali, permettendo una maggiore precisione e personalizzazione degli interventi, in particolar modo per quegli edifici che presentano vincoli particolari, come gli edifici storici o sottoposti a tutela.

La riqualificazione energetica degli edifici con considerevoli vincoli storici e architettonici rappresenta una sfida complessa ma possibile. Questo tipo di edifici, spesso protetti da normative che ne preservano l'integrità storica e culturale, richiede un approccio delicato e innovativo per migliorare l'efficienza energetica senza comprometterne il loro valore storico e architettonico. Il Caso Studio presentato all'interno del paper costituisce un esempio

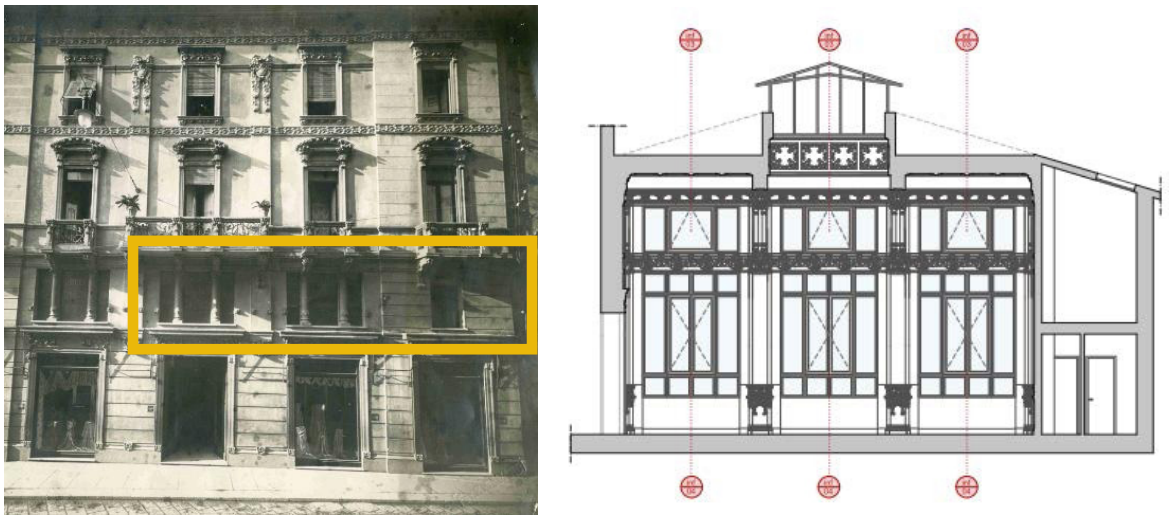
di riqualificazione di un edi-ficio di particolare valore storico-architettonico.

## 1. Descrizione del progetto

Uno dei casi studio di P2Endure si trova a Reggio Emilia, in Italia. Questo Caso Studio è particolarmente significativo perché rappresenta un esempio concreto di come le soluzioni di P2Endure possono essere applicate in un contesto reale. Il Caso Studio ha coinvolto un edificio successivamente utilizzato come ufficio e si-tuato nel centro storico di Reggio Emilia.

Il caso studio è un edificio storico, non abitato dalla fine del 900, vincolato dalla Sovrintendenza per via del suo valore storico e artistico. L'obiettivo del progetto è quello di riqualificare energeticamente l'edificio, preservando i suoi elementi storici, rispettando i vincoli della Soprintendenza.

Fig. 1  
Prospetto e sezione del Caso Studio



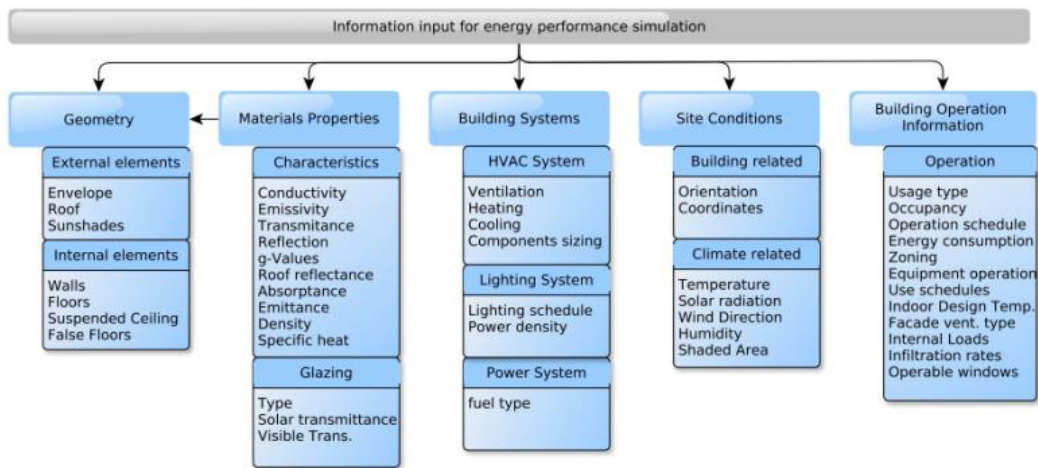
## 2. Generazione del modello digitale

La prima attività avviata per la definizione del modello digitale del Caso Studio è stata la raccolta dei dati. La precisione del modello è infatti in stretta correlazione con il dettaglio delle informazioni raccolte. In particolare, si possono individuare i seguenti 3 gradi di approfondimento:

- 1° grado: conoscenza generale dell'edificio e delle attività che vi si svolgono;
- 2° grado: un maggior livello di dettaglio del modello è raggiungibile tramite un audit sottoposto agli abitanti al fine di determinarne i profili di consumo energetico e i pattern di utilizzo in relazione alle attività svolte in ogni locale. La conoscenza della stratigrafia dei singoli componenti, nonché delle caratteristiche tecniche delle apparecchiature installate permettono di definire con precisione il fabbisogno energetico di ciascun dispositivo;
- 3° grado: il maggior livello di conoscenza dell'edificio si raggiunge attraverso la creazione di un modello energetico digitale, il quale simula il comportamento dell'edificio reale in ogni condizione climatica. Tale grado di dettaglio può essere raggiunto implementando, le

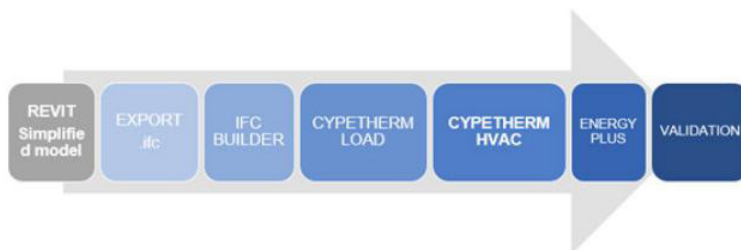
informazioni raccolte nelle fasi precedenti all'interno del modello digitale. Per il Caso Studio oggetto del presente articolo è stato raggiunto il grado più elevato di conoscenza dell'edificio: sono infatti state raccolte informazioni riguardanti sia gli impianti installati sia le partizioni orizzontali e verticali, opache e trasparenti, preesistenti. Essendo il Caso Studio non abitato da diversi anni, non è stato possibile utilizzare una base storica relativa ai dati di consumo. Per superare questa criticità, il fabbisogno energetico è stato stimato utilizzando i dati relativi a edifici che presentavano le medesime caratteristiche, nonché le informazioni ricavabili dalla Certificazione Energetica disponibile. In Fig. 2 un riepilogo dei dati necessari per una precisa definizione del modello digitale.

Fig. 2  
Dati necessari per la definizione del modello digitale



La definizione del modello digitale può avvenire tramite l'utilizzo di diversi software disponibili sul mercato i quali permettono l'esportazione del modello digitale (BIM) e la creazione del modello energetico (BEM) tramite l'utilizzo di file in formati specifici (.ifc, .idf, .gbXML).

Fig. 3  
Esempio di processo per la creazione del modello energetico digitale



Per il presente Caso Studio il modello energetico è stato realizzato tramite la piattaforma Design Builder, interfaccia grafica di Energy Plus. L'utilizzo di tale software ha permesso di limitare il numero di esportazioni/importazioni e i conseguenti errori geometrici che derivano dall'utilizzo di piattaforme diverse. Il processo è stato quindi notevolmente semplificato rispetto a quanto riportato in Fig.3.

Una volta realizzato il modello geometrico, comprensivi dei dati stratigrafici dei singoli componenti, sono state inserite le informazioni riguardanti l'impiantistica presente e i relativi profili di utilizzo.

E' stata successivamente condotta una simulazione energetica per verificare la corrispondenza dei risultati ottenuti con i dati disponibili; il modello risulta "Validato" nel caso in cui i due valori risultino analoghi (considerando una tolleranza pari al 2%).

Una volta validato, il modello è stato utilizzato per stimare l'efficacia degli interventi di riqualificazione previsti.

### 3. Stima della efficacia degli interventi di riqualificazione

Considerando il valore storico dell'edificio, gli interventi di riqualificazione ammissibili sono stati selezionati con l'obiettivo di preservare le caratteristiche esistenti. Al fine di stimare l'efficacia dei singoli interventi, sono stati raccolti e inseriti i dati specifici all'interno del modello digitale ed è stata elaborata una simulazione energetica.

E' stato quindi possibile valutare, in una fase preliminare rispetto all'inizio dei lavori, i vantaggi (in termini sia energetici che economici) della applicazione delle singole tecnologie. Sono state successivamente elaborate simulazioni energetiche che comprendessero combinazioni diverse di tali tecnologie, attraverso un processo iterativo che ha permesso di individuare il set di interventi che permettesse di raggiungere i benefici maggiori.

Fig. 4  
processo iterativo di individuazione interventi

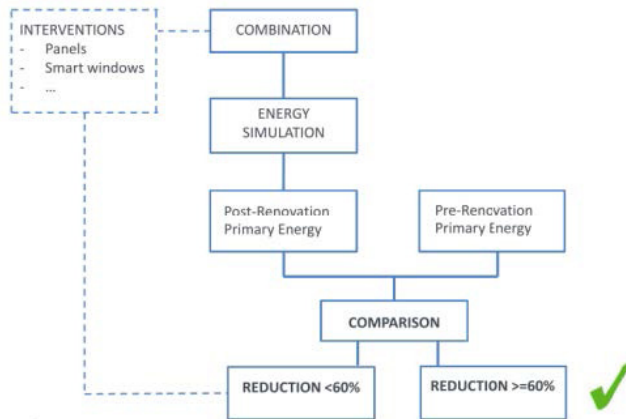


Fig. 5  
individuazione interventi e stima efficacia

<b>FINAL COMBINATION</b>	<ul style="list-style-type: none"> <li>▪ P2ENDURE TECHNOLOGIES Smart Windows BGTEC</li> <li>▪ OTHER TECHNOLOGIES Internal insulation + Consensing boiler + LED lamps</li> </ul>
<b>ENERGY CONSUMPTION POST-RENOVATION</b>	<ul style="list-style-type: none"> <li>▪ PRIMARY ENERGY CONSUMPTION IN PRE-RENOVATION SCENARIO [kWh/m2y] 161 kWh/m2y</li> <li>▪ PRIMARY ENERGY CONSUMPTION IN POST-RENOVATION SCENARIO [kWh/m2y] 64 kWh/m2y</li> <li>▪ SAVING [%] 60.20%</li> </ul>

In Fig. 5 è dettagliato il set di tecnologie la cui applicazione permette di raggiungere il maggior beneficio, in termini di fabbisogno di energia primaria.

#### 4. Monitoraggio

A seguito della realizzazione degli interventi sono stati raccolti i dati relativi ai consumi energetici e, considerando il fattore di conversione da energia elettrica a energia primaria specifico, è stato calcolato il fabbisogno di energia primaria del Caso Studio.

Questi dati sono stati confrontati con i consumi pre-intervento all'interno della Tab. 1 (in corsivo i dati pre-intervento): in rosso è calcolata la riduzione di energia primaria rispetto all'anno 2018, preso come riferimento.

Tab. 1  
Dati di consumo di energia primaria (espresso in kWh) per ogni anno di utilizzo

	2018	2019	2020	2021	2022	2023
<b>gen</b>	<i>15934.67</i>	222.64	1592.36	6030.64	5055.38	4169.66
<b>feb</b>	<i>10862.12</i>	212.96	7126.90	2879.80	3446.08	3547.72
<b>mar</b>	<i>11342.68</i>	118.58	1311.64	6030.64	3598.54	2848.34
<b>apr</b>	<i>5896.36</i>	111.32	1103.52	5834.62	1870.66	1846.46
<b>mag</b>	<i>4675.90</i>	1057.54	1134.98	2571.25	1483.46	1437.48
<b>giu</b>	<i>4508.09</i>	1045.44	1314.06	2571.25	1430.22	1439.90
<b>lug</b>	<i>5652.27</i>	1064.80	1113.20	3794.56	1793.22	1769.02
<b>ago</b>	<i>5766.69</i>	1384.24	1355.20	1679.48	1829.52	1580.26
<b>set</b>	<i>4492.83</i>	1202.74	1478.62	1277.76	1425.38	1376.98
<b>ott</b>	<i>4302.13</i>	1214.84	2013.44	1967.46	1364.88	1551.22
<b>nov</b>	<i>9664.54</i>	762.30	2376.44	2490.18	3066.14	2891.90
<b>dic</b>	<i>13142.86</i>	762.30	4568.96	4322.12	4169.66	4126.10
<b>TOT</b>	<b>96241.163</b>	<b>9159.7</b>	<b>26489.32</b>	<b>41449.76</b>	<b>30533.14</b>	<b>28585.04</b>
		-90%	-72%	-57%	-68%	-70%

Una ulteriore implementazione è stata valutata al fine di provvedere al residuo fabbisogno energetico tramite fonti rinnovabili.

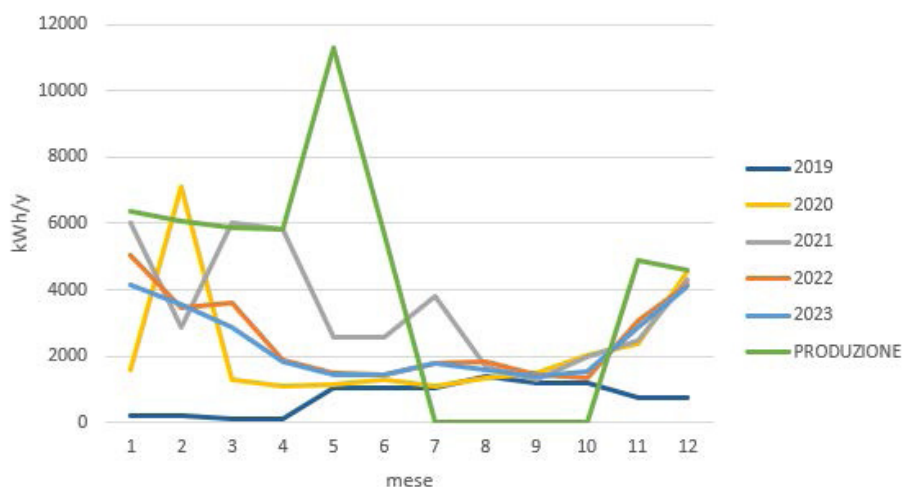
La società titolare del Caso Studio, infatti, dispone di un mini-impianto idroelettrico su canale situato nella medesima provincia.

L'impianto idroelettrico e l'edificio, pertanto, si possono considerare virtualmente connessi, prevedendo l'utilizzo, da parte del Caso Studio, dell'energia primaria prodotta dall'impianto idroelettrico.

Si prende pertanto in esame l'energia primaria prodotta dall'impianto idroelettrico in un anno medio: la produzione di energia elettrica risulta non costante nell'anno solare, con picchi nei periodi primaverili e arresto della produzione nei mesi più torridi.

Il bilancio energetico viene quindi verificato su base annuale. In Fig. 6 si riassume la verifica condotta al riguardo, la quale conferma che il fabbisogno annuo residuo del Caso Studio viene interamente soddisfatto dal vicino impianto idroelettrico.

Fig. 6  
Confronto flussi di energia primaria



## 5. Conclusioni

A seguito del progetto condotto, si desume che la riqualificazione energetica degli edifici storici è possibile attraverso un approccio integrato che rispetta e valorizza il patrimonio culturale.

Un'analisi preventiva dettagliata e mirata permette di identificare le tecnologie più efficaci e compatibili con le caratteristiche uniche degli edifici storici, senza compromettere il loro valore storico e architettonico. Questo approccio non solo garantisce il successo della riqualificazione energetica, ma preserva anche il valore culturale degli edifici, assicurando interventi sostenibili e rispettosi del patrimonio esistente. Il BIM si delinea come strumento potente e versatile che offre numerosi vantaggi nel processo di riqualificazione energetica, in particolare degli edifici storici: permette infatti di affrontare le complessità legate alla preservazione del patrimonio culturale, migliorando al contempo l'efficienza energetica e la sostenibilità. Grazie alla tecnologia BIM, è possibile pianificare, eseguire e gestire interventi di riqualificazione in modo più accurato, coordinato ed efficace, rispettando i vincoli storici e architettonici degli edifici.

Inoltre, il processo di monitoraggio del Caso Studio ha consentito di verificare che la generazione diffusa di energia elettrica rappresenta una risposta chiave ai nuovi fabbisogni energetici, offrendo numerosi vantaggi in termini di efficienza, sostenibilità, affidabilità e coinvolgimento delle comunità locali. Il caso studio rappresenta perfettamente questo schema, riprendendo quello che è lo schema tipico delle Comunità Energetiche Rinnovabili.

## Autori

Giacomo Bizzarri [giacomo.bizzarri@unife.it](mailto:giacomo.bizzarri@unife.it)

Ricercatore di Fisica Tecnica Industriale e Professore Aggregato di "Fisica Tecnica Ambientale 1" presso la Facoltà di Architettura, Università degli Studi di Ferrara. È autore di oltre quaranta articoli scientifici, pubblicati su riviste nazionali e internazionali; ha inoltre partecipato come Speaker a oltre undici Congressi Internazionali.

**Enrica Boldrin** [bldnrc@unife.it](mailto:bldnrc@unife.it)

Laureata in Architettura presso la facoltà di Architettura dell'Università di Ferrara, ho sempre lavorato nell'ambito della riqualificazione energetica, con un particolare interesse per l'applicazione delle fonti rinnovabili e alle riqualificazioni energetiche. Il Dottorato in Architettura e Pianificazione Urbana presso l'Università di Ferrara ha permesso di approfondire la ricerca nel settore energetico.

**Laura Ferrari** [laura01.ferrari@unife.it](mailto:laura01.ferrari@unife.it)

Laureata in Ingegnerie Edile e Architettura presso la facoltà di Ingegneria dell'Università di Bologna, ha sempre lavorato nell'ambito della riqualificazione energetica, con un particolare interesse per l'applicazione delle fonti rinnovabili per soddisfare i fabbisogni energetici. Il Dottorato in Architettura e Pianificazione Urbana presso l'Università di Ferrara ha permesso di approfondire la ricerca nel settore e di studiare l'applicazione delle più recenti infrastrutture digitali.

# Data, Reduced Order Models and Artificial Intelligence for Digital Twin

Gianluigi Rozza<sup>1,2</sup>, Laura Meneghetti<sup>1</sup>, Andrea Martini<sup>1,2</sup>

<sup>1</sup>SISSA mathLab , Trieste, Italy, <sup>2</sup>FAST Computing, Trieste

**Abstract.** An increasing number of disruptive innovations with high economic and social impact are characterizing the digitalization processes occurring in the last years. Despite the benefits of these developments, their speed and extent are limited by the available simulation technologies to handle complex models. The key idea of Digital Twin (DT) lies exactly in the creation of a digital counterpart of the physical asset able to replicate its behaviour. The necessity of a continuous and fast interaction between the physical and digital environments through background simulations has led to the introduction of Reduced Order Modeling (ROM) and Artificial Intelligence (AI) as crucial technologies to speed up the model execution, while maintaining good level of accuracy but with a reduced number of degrees of freedom. In this way, DT can be applied to new complex emerging fields, providing also more accurate information on the real counterpart, thanks to the presence of these techniques.

**Keywords.** Digital Twin, Reduced Order Model, Data Analysis, Artificial Intelligence, Real Time Computing

## Introduction

Reduced Order Modelling is a quickly emerging field in applied mathematics, computational science and engineering for speeding up Numerical Simulations. The growing demand of efficient computational tools and real time computations, coupled with the presence of parametric formulations and uncertainty quantification in the models under consideration have led to the necessity of a computational collaboration between High Performance Computing(HPC) and Reduced Order Methods(ROMs)[14]. The former deals with high order models, characterized by very expensive and time demanding simulations requiring HPC facilities. ROMs enable to overcome these difficulties by constructing a reduced version of the full model, faster than the previous one. In this way, real-time input-output evaluations are possible without requiring demanding resources.

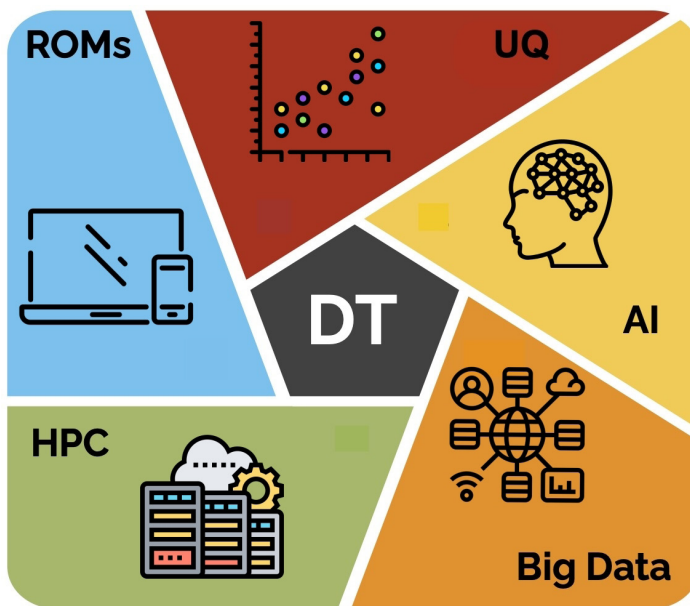
Artificial Intelligence (AI), Deep Learning (DL) and Machine Learning (ML) are revolutionizing the field of Computational Science bringing high generalization capability. Neural networks, with their capacity to learn and approximate complex patterns, have emerged as powerful tool in several contexts. Therefore, integrating neural networks into ROMs offers a powerful approach to approximate complex systems efficiently and accurately. DL plays an important role also in managing the huge amount of data at disposal nowadays. This vast amount of data, constantly generated by several sources , requires advanced tools and techniques for storage, processing, and analysis to uncover valuable insights and drive decision-making. The use of ML techniques coupled with HPC allow the mana-

gement of Big Data and a real-time analysis of the continuous flow of such data coming from various sources such as simulations, sensors, etc. They also can ensure an estimation of the uncertainty on such data to ensure data veracity and the capacity of the extraction of knowledge and significant patterns from the available data.

Digital Twins (DTs) represent thus a powerful tool for integrating ROMs, HPC, AI, Big Data and Uncertainty Quantification. By creating a virtual replica of physical systems, DTs enable real-time monitoring and simulations, handled through the application of HPC and ROMs. AI can then analyze vast amounts of data for predictive maintenance and optimization. The synergy of these technologies with Big Data ensures more accurate and efficient decision-making, enhancing the performance and reliability of systems across various applications.

### 1. An integrated paradigm: Digital Twin

Fig. 1  
An integrated  
paradigm: Digital  
Twin



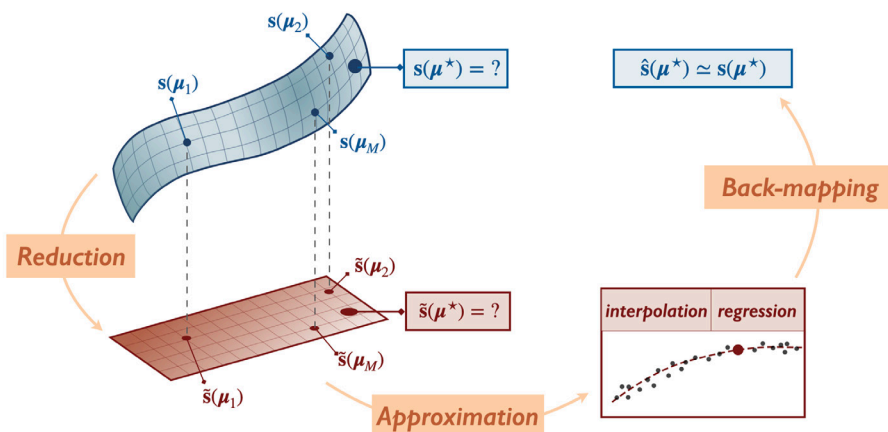
A Digital Twin (DT) represents a virtual replica of an object or phenomena, which can be used to gain improvements in the physical part. This is obtained through an automatic flow of data from the physical object to the digital one, making DT a fundamental support tool for the entire life-cycle of an object.

As represented in Figure 1, DT represents a paradigm integrating several elements:

1. **Big Data:** Data are the basis for the construction of such models. Nowadays, IoT devices represent a great source of data due to their increasing presence in everyday objects, factories, etc.
2. **High Performance Computing (HPC):** This huge amount of data needs to be handled through suitable infrastructures able to capture the underlying structure and information

contained inside them. This amount of data also represents the input for some mathematical model describing the problem under consideration. These models are often characterized by a high number of parameters due to the complexity of the modeled problem. HPC is thus fundamental in storing Big Data and running complex simulations.

Fig. 2  
Schematic representation of the reduction process for ROMs



3. Reduced Order Models (ROMs): Due to the complexity (high number of parameters) of models, simulations are slow and very expensive from the computational point of view. To be applied in an industrial environment, simulations need to be real-time and with a low computational cost. ROMs represent thus an approach to handle this problem, providing models reduced in the number of parameters but with high accuracy. The goal of these methods is to detect the most important direction and information in the parameter space, to perform a reduction based on them. An approximated solution can then be constructed through an interpolation/regression process.

4. Artificial Intelligence (AI): AI techniques, such as machine learning and neural networks, enhance simulations by learning complex patterns from data, improving accuracy and efficiency. These methods can also be integrated in ROMs to enable better approximations of complex systems.

5. Uncertainty Quantification (UQ): UQ refers to the process of assessing and managing uncertainties that arise from both data and the mathematical models used to interpret that data. It is thus essential for improving the reliability and applicability of models in several fields.

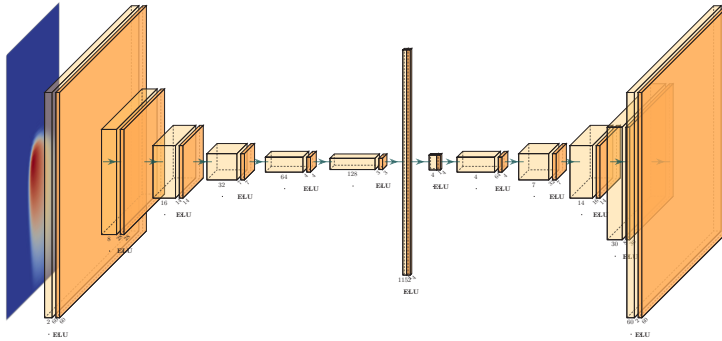
## 2. Reduced Order Models enhanced by Deep Learning and Scientific Machine Learning

Artificial Intelligence represents a tool increasingly integrated in several fields. In particular, it can be employed to enhance classical (data-driven) ROM techniques for Computational Fluid Dynamics. Some examples include the following methods:

- Approximation in Reduced Order Model: ANNs are introduced to extract information

coming from data and to approximate the functional relationship between input parameters and output solutions. POD-NN and AE-NN represent possible techniques developed for this purpose [1, 2, 3, 4, 8]. Neural Networks can thus be adopted to improve the accuracy of a POD-based model by improving generalization and providing a multi-fidelity perspective.

Fig. 3  
A convolutional  
Auto-encoder



- Auto-encoders for dimensionality reduction and manifold learning: Autoencoders are included in data-driven methods to learn a latent representation of the dynamical system. To include the knowledge about the physics of the model, we can couple a non-linear manifold method [6] with hyper-reduction achieved through reduced over-collocation and teacher-student training of a reduced decoder [5].
- Reduction in wide parameter space by means of deep learning parameter domain decomposition: A Deep Learning framework can be developed to generalize the evolution of a system over initial conditions in an extended parameter space. In [7] a two stage memory aware ANNs model order reduction approach has been constructed to deal with this problem. Curse of dimensionality is initially tackled by partitioning the parameter space and averaging ROM solutions. Then Long-short term memory network (LSTM) are coupled with Convolutional Networks (C-LSTM) for extracting temporal correlations.
- Generative models to quantify model uncertainty: As described in the previous section, UQ represents an important element to keep into account. In [13], an approach integrating generative modeling with UQ is presented, where the former is employed to learn probability distributions on the data and to capture the intrinsic characteristics of the original ones. A priori UQ quantification can thus be done exploiting this probabilistic approach.

## 2.1 Machine Learning pipeline

Given a problem (e.g. fluid dynamics, stochastic PDEs), a pipeline can be defined to include Machine Learning in the solution of the corresponding PDEs. The first step coincide with the generation of data for the problem under consideration. Hence, high fidelity simulations can be performed or scattered data from the domain can be determined. Once the problem and the data have been determined, a ML model can be built, using for example

Neural Networks, POD with Interpolation, Neural Operators, etc.. The techniques listed above represent possible solutions for the construction of ML models. The last step coincides then with the optimization of the model using, for example, Supervised Learning, Physics-Informed losses and the gradient descent method.

An example of integration of Deep Learning into PDEs is represented by Physics-Informed Neural Networks (PINNs) [9,10, 11,12], an optimization technique to compute solution of differential equation using Neural Networks

In this framework, PDEs are solved by including physical laws and symmetries in Neural Networks. PINNs represent thus an approach that balances data and physical knowledge, coming from ROMs simulations, in order to build a truthful and reliable ML model.

Fig. 4  
Machine Learning pipeline

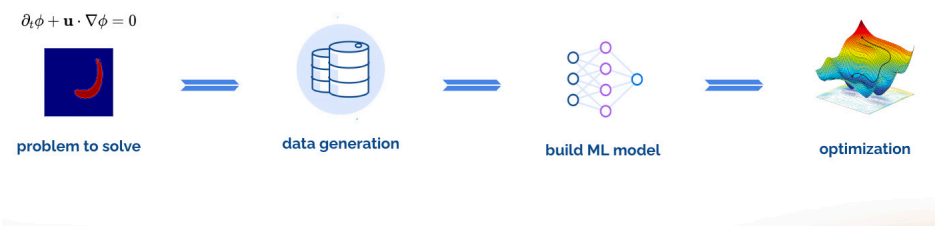
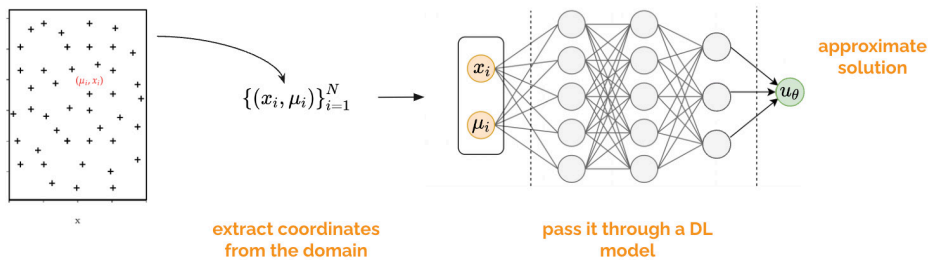


Fig. 5  
PINNs pipeline



### 3. Applications of Digital Twin

As explained in the previous sections, a digital twin is the virtual replica of a physical asset. It means that it is possible to realize a digital twin in very different fields. It could be a DT of product or processes which means, for instance, a manufacturing asset or manufacturing plant; but it can be translated into a DT of an environment or entire processes. We developed three Digital Twins, more specifically three Live Demos, related to the Competence center Industry 4.0 of the north-east of Italy, SMACT Odyssea. Those are three useful and tangible examples to understand the potential of the tool: one is a Smart Hybrid Energy Management System with Wartsila connected to hybrid engines of vessels; the second one is a Heavy-Duty Machining Processes Optimisation with Danieli Automation and the last one, still under development, is related to an aspiring hood with Electrolux. We do mention also the activities of Spoke 9 of iNEST, Interconnected Innovation Ecosystem of North-East of Italy dedicated to computational technologies for digital twins: including infrastructures, services, networks, communities as well as medicine. The main concept which stands behind this innovative tool is the flux of information coming from

the real asset going to the virtual replica and come back to the real one: it is crucial to obtain real time information necessary to take appropriate actions in case of necessity.

## 4. Conclusions

Digital twins are revolutionizing industries by offering virtual replicas that simulate physical assets, processes, or systems in real-time. This technology enables proactive maintenance, performance optimization, and efficient resource management. Data, Reduced Order Modeling, Control, Optimization, Machine Learning, Artificial Intelligence and Uncertainty Quantification play a key role in the definition of DTs, leading to the need of a better integration to create this new parametrized, reduced and coupled paradigm. Applied Mathematics is thus fundamental in the advances of Science and Engineering, becoming a propeller for methodological innovation and technology transfer.

## Bibliography

- [1] San, O., Maulik, R., & Ahmed, M. (2019). An artificial neural network framework for reduced order modeling of transient flows. *Communications in Nonlinear Science and Numerical Simulation*, 77, 271-287.
- [2] Hesthaven, J. S., & Ubbiali, S. (2018). Non-intrusive reduced order modeling of nonlinear problems using neural networks. *Journal of Computational Physics*, 363, 55-78.
- [3] Papapicco, D., Demo, N., Girfoglio, M., Stabile, G., & Rozza, G. (2022). The neural network shifted-proper orthogonal decomposition: a machine learning approach for nonlinear reduction of hyperbolic equations. *Computer Methods in Applied Mechanics and Engineering*, 392, 114687.
- [4] Shah, N. V., Girfoglio, M., Quintela, P., Rozza, G., Lengomin, A., Ballarin, F., & Barral, P. (2022). Finite element based Model Order Reduction for parametrized one-way coupled steady state linear thermo-mechanical problems. *Finite Elements in Analysis and Design*, 212, 103837.
- [5] Romor, F., Stabile, G. & Rozza, G. Non-linear Manifold Reduced-Order Models with Convolutional Autoencoders and Reduced Over-Collocation Method. *J Sci Comput* 94, 74 (2023). <https://doi.org/10.1007/s10915-023-02128-2>
- [6] Lee, K., & Carlberg, K. T. (2021, May). Deep conservation: A latent-dynamics model for exact satisfaction of physical conservation laws. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 1, pp. 277-285).
- [7] Gonnella, I. C., Hess, M. W., Stabile, G., & Rozza, G. (2023). A two-stage deep learning architecture for model reduction of parametric time-dependent problems. *Computers & Mathematics with Applications*, 149, 115-127.
- [8] Demo, N., Tezzele, M. & Rozza, G. (2023). A DeepONet multi-fidelity approach for residual learning in reduced order modeling. *Adv. Model. and Simul. in Eng. Sci.* 10, 12 . <https://doi.org/10.1186/s40323-023-00249-9>
- [9] Raissi, Maziar, Paris Perdikaris, and George E. Karniadakis. "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations." *Journal of Computational physics* 378 (2019):

686-707.

[10] Cuomo, Salvatore, et al. "Scientific machine learning through physics-informed neural networks: Where we are and what's next." *Journal of Scientific Computing* 92.3 (2022): 88.

[11] Demo, N., Strazzullo, M., & Rozza, G. (2023). An extended physics informed neural network for preliminary analysis of parametric optimal control problems. *Computers & Mathematics with Applications*, 143, 383-396.

[12] Coscia, D., Ivagnes, A., Demo, N., & Rozza, G. (2023). Physics-Informed Neural networks for Advanced modeling. *Journal of Open Source Software*, 8(87), 5352.

[13] Coscia, D., Demo, N., & Rozza, G. (2024). Generative Adversarial Reduced Order Modelling. *Nature Scientific Report*.

[14] Rozza, G., Stabile, G., & Ballarin, F. (Eds.). (2022). *Advanced reduced order methods and applications in computational fluid dynamics*. Society for Industrial and Applied Mathematics.

## Authors



**Gianluigi Rozza** [gianluigi.rozza@sissa.it](mailto:gianluigi.rozza@sissa.it)

Gianluigi Rozza is a Full Professor at SISSA and also coordinator of SISSA mathLab group, an applied mathematical research group devoted to develop mathematical modeling solutions to be applied to industrial, medical and environmental entities. He has a PhD - Doctoral School in Applied Mathematics (Numerical Analysis, Computational Engineering) from EPFL- Ecole Polytechnique Fédérale de Lausanne (Switzerland) and he was a postdoctoral research fellow at MIT- Massachusetts Institute of Technology, Boston MA (USA) at the Department of Mechanical Engineering and Center for Computational Engineering.

**Laura Menghetti** [laura.meneghetti@sissa.it](mailto:laura.meneghetti@sissa.it)

Laura Meneghetti is a Research Fellow at SISSA in the iNEST project. Previously, she studied Mathematics at the University of Padua, then moved to Trieste to obtain her PhD in Mathematical Analysis, Models, and Applications at SISSA. During her PhD, she specialized in studying neural networks for the problem of Object Recognition. In particular, she developed a mathematical method to reduce the space occupied by such networks to be integrated into embedded vision systems. Then she has been involved in other projects that involve Machine Learning and Data Science techniques.



**Andrea Martini** [andrea.martini@fastcomputing.net](mailto:andrea.martini@fastcomputing.net)

Andrea Martini is a researcher within SISSA mathLab in charge of Valorization and Project Management. Since a couple of years he is CEO of FAST Computing, an official SISSA startup with the focus of implementing innovative mathematical modeling to boost and optimize industrial products and processes. He has a master degree in Law and he got a Master in Business Administration in International Business obtained at MIB Trieste School of Management.

# Una infrastruttura Cloud federata per ICSC e TeRABIT

Claudio Grandi

INFN-Bologna

**Abstract.** ICSC, il Centro Nazionale di Ricerca in High Performance Computing, Big Data e Quantum Computing, e TeRABIT, il progetto sulle infrastrutture di ricerca che sta potenziando le infrastrutture digitali di GARR, CINECA e INFN in totale sinergia con ICSC, sono ormai a metà del loro percorso. Anche se la maggior parte dei fondi è dedicata al potenziamento delle infrastrutture e all'acquisizione di risorse, una parte considerevole del lavoro è dedicata alla creazione di un sistema capace di federare le risorse gestite da entità diverse. In questo contributo si descrivono le attività per creare una infrastruttura Cloud federata per ICSC e TeRABIT.

**Keywords.** HPC, HTC, Network, Cloud, Federation

## Introduzione

Il Centro Nazionale di Ricerca in HPC, Big Data e Quantum Computing (ICSC)<sup>1</sup> è stato finanziato nell'iniziativa From Research to Business con 320 milioni di euro. Il progetto Terabit Network for Research and Academic Big Data in Italy (TeRABIT)<sup>2</sup> è stato finanziato con 41 milioni di euro nell'iniziativa Research Infrastructures. 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 digitali di GARR, CINECA e INFN in totale sinergia con ICSC. Una descrizione più ampia dei progetti si può trovare anche in [1].

Nel documento si descrivono le attività legate all'implementazione di una federazione cloud nazionale in grado di consentire un accesso trasparente all'infrastruttura e un utilizzo efficiente delle risorse. In particolare, si descrive il sistema di test in fase di realizzazione con funzionalità ancora limitate ma che costituisce il primo passo verso il sistema di produzione.

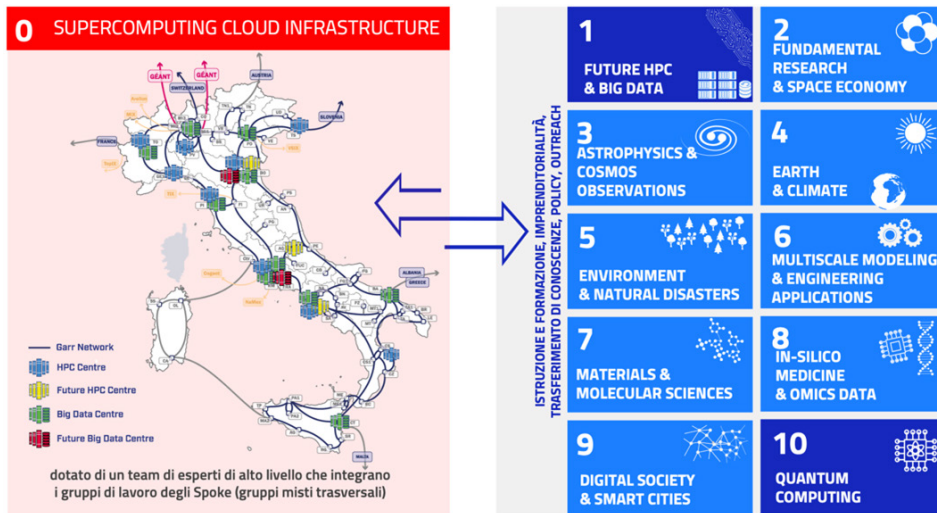
## 1. Organizzazione del progetto ICSC

ICSC è iniziato nel settembre 2022 e la fase di implementazione 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. I partner del progetto sono università, enti di ricerca e operatori pubblici e privati.

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

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

Fig. 1  
La descrizione  
degli spoke di  
ICSC



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. L'implementazione della cloud federata è parte delle attività di Spoke-0.

## 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: 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 le stesse dello Spoke-0 di ICSC. In effetti i due progetti sono costruiti per essere totalmente sinergici.

## 3. L'infrastruttura Cloud Federata

L'infrastruttura nazionale sarà composta da centri di calcolo HPC di diverse dimensioni, HTC e Big Data collegati da una rete ad alte prestazioni. Il modello è incentrato sui dati, con nodi storage interconnessi tramite una rete a larga banda e nodi di calcolo eterogenei che accedono ai dati in modo efficiente indipendentemente dalla loro locazione.

I principi su cui si basa il processo federativo sono: l'inclusività, attraverso una federazione "leggera" e l'uso di standard, come ad esempio l'uso di federazioni d'identità; la facilità d'uso, grazie ad un orchestratore PaaS e dashboard costruite sulle esigenze degli utenti; la flessibilità, grazie a meccanismi ibridi di allocazione delle risorse che consentano di mantenere efficienza ed efficacia nell'uso delle risorse, consentendo al tempo stesso la trasparenza e la possibilità di indirizzare casi d'uso specifici.

Il modello di gestione delle risorse è descritto in figura 2. Nello strato più basso ci sono le risorse, che sono gestite dai rispettivi enti in modo autonomo. Nel secondo layer ci sono servizi che implementano funzionalità di base, quali il trasferimento dati o l'accesso alle risorse di calcolo, comuni alla maggior parte delle applicazioni. Nel terzo strato ci sono

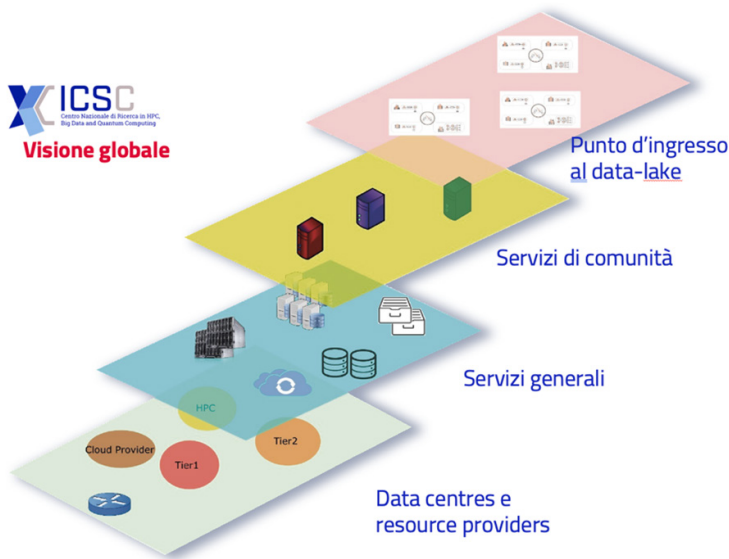
servizi di comunità, sviluppati dagli esperti dei diversi ambiti applicativi per implementare funzionalità specifiche. Lo strato più alto è il punto di ingresso per le diverse comunità applicative.

Ci sono diverse sfide che devono essere affrontate per raggiungere il risultato.

Per implementare un processo di autenticazione e autorizzazione federata è necessario costruire un meccanismo di fiducia, di trust, fra le istituzioni partecipanti, che sia compatibile con i requisiti operativi delle singole infrastrutture. Inoltre, è necessario definire meccanismi standard per comunicare il livello di garanzia, di assurance, delle singole operazioni, ad esempio il REFEDS Assurance Framework<sup>3</sup>.

Per l'accesso alle risorse di calcolo è necessario usare soluzioni standard, quali OpenStack<sup>4</sup>, Kubernetes<sup>5</sup>, che siano supportate dagli strumenti della federazione. Analogamente, per l'accesso ai sistemi di archiviazione dati bisogna essere compatibili con diverse tipologie di sistemi, come ad esempio Amazon S3<sup>6</sup> o OpenStack SWIFT<sup>7</sup>.

Fig. 2  
Il modello a strati  
dell'infrastruttura  
ICSC



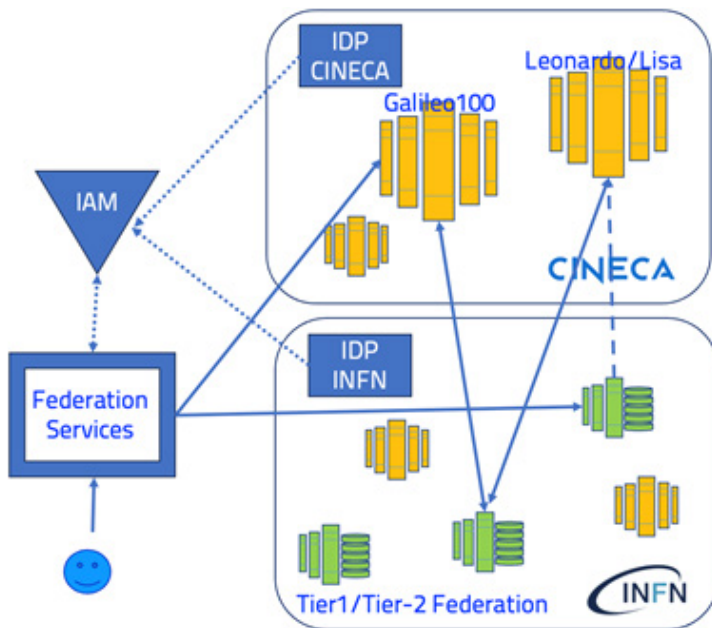
L'aspetto più sfidante del processo è che tutto questo deve essere fatto su sistemi in produzione senza alterarne l'operatività.

#### 4. Il Proof of Concept della federazione

Nel giugno 2024 è in attivazione un dimostratore (Proof of Concept, PoC) della federazione, che pur con soluzioni temporanee, consente di testare le componenti principali del sistema: accesso e gestione delle identità, federazione delle risorse di calcolo, federazione dei dati. In questo primo PoC, sono state federate le risorse di CINECA e INFN; l'uso della rete GARR è totalmente trasparente.

Come mostrato in figura 3, l'utente si autentica tramite i servizi della federazione usando Indigo-IAM<sup>8</sup>. IAM è connesso a due sistemi di gestione delle identità (Identity Provider):

Fig. 3  
Il Proof of Concept  
della federazione  
cloud di giugno  
2024



quello di CINECA e quello dell'INFN. Utenti registrati in uno dei due sistemi possono essere autorizzati all'uso delle risorse. Una delle limitazioni correnti è che per accedere ai sistemi CINECA l'utente deve essere registrato sui sistemi CINECA, mentre l'INFN consente l'accesso anche a utenti registrati solo al CINECA, ovviamente se autorizzati. Una volta autenticato e autorizzato, l'utente potrà utilizzare risorse dei sistemi cloud e storage di CINECA e INFN. Inoltre, tramite un processo di offloading, potrà accedere, se autorizzato, alle code batch del sistema HPC Leonardo del CINECA, pur non essendo presente una interfaccia Cloud nativa.

Questo sistema evolverà in modo da includere altri Identity Provider, quali ad esempio quelli che aderiscono alla federazione IDEM<sup>9</sup>, e altre risorse.

## 5. Conclusioni

A circa metà della durata dei progetti ICSC e TeRABIT, si sta procedendo nella costruzione della Cloud federata nazionale, con l'intento di fornire alla comunità della ricerca nazionale un accesso trasparente ma efficiente alle risorse di calcolo. Oltre che tecnologico, il problema è di tipo sociologico e organizzativo, perché per raggiungere lo scopo è necessario individuare e concordare strumenti e processi condivisi, che consentano agli enti che gestiscono le risorse di continuare le attività di produzione senza alterarne la funzionalità. L'approccio scelto è quello della federazione "leggera", tramite uno strato di orchestrazione PaaS, l'adozione di standard e lo sviluppo di componenti che consentano una flessibilità d'uso delle risorse.

## Riferimenti bibliografici

[1] Grandi C., Campanella M (2024), I progetti ICSC, il Centro Nazionale HPC, Big Data e Quantum Computing e TeRABIT, Conferenza GARR 2023 - Saperi interconnessi - Se-

lected Papers Firenze 14-16 giugno 2023 - ISBN: 978-88-946629-2-4, DOI: 10.26314/GARR-Conf23-proceedings-17

## Ringraziamenti

Le attività sono finanziate dai progetti PNRR, ICSC (M4, C2, I1.4, CP CN00000013, CUP I53C21000340006) e TeRABIT (M4, C2, I3.1, CP IR0000022, CUP I53C21000370006).

## Autori

Claudio Grandi [Claudio.Grandi@bo.infn.it](mailto: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; è co-leader di Spoke-0 di ICSC e membro del Comitato Scientifico di TeRABIT.

## Notes

<sup>1</sup>ICSC: <https://www.supercomputing-icsc.it/en/icsc-home/>

<sup>2</sup>TeRABIT: <https://www.terabit-project.it/it/>

<sup>3</sup>REFEDS Assurance Framework: <https://refeds.org/assurance>

<sup>4</sup>OpenStack: <https://www.openstack.org/>

<sup>5</sup>Kubernetes: <https://kubernetes.io/it/>

<sup>6</sup>Amazon S3: <https://aws.amazon.com/it/s3/>

<sup>7</sup>OpenStack SWIFT: <https://docs.openstack.org/swift/latest/>

<sup>8</sup>INDIGO-IAM: <https://github.com/indigo-iam/iam>

<sup>9</sup>IDEM: <https://www.idem.garr.it/>

# Progetti di comunicazione Quantistica nella Regione Friuli Venezia Giulia

Angelo Bassi<sup>1,2</sup> e Giorgio Giorgetti<sup>2</sup>

<sup>1</sup>INFN sezione di Trieste, <sup>2</sup>Università degli Studi di Trieste

**Abstract.** I progetti "Quantum FVG" e "QuFree" finanziati dalla Regione Friuli Venezia Giulia (FVG) mirano a realizzare una infrastruttura regionale permanente per la comunicazione sicura tramite la Distribuzione Quantistica delle Chiavi (QKD). Il progetto "Quantum FVG" ha stabilito un laboratorio all'avanguardia di comunicazione quantistica presso il CNR-INO di Trieste, promuovendo la ricerca, l'istruzione e lo sviluppo dell'infrastruttura quantistica regionale. Complementariamente, il progetto "QuFree" esplora la comunicazione quantistica nello spazio libero. Le prospettive future includono l'espansione dell'infrastruttura quantistica per ampliare ulteriormente le comunicazioni sicure nella Regione, integrandosi con reti nazionali ed europee e sviluppando applicazioni in vari settori, in particolare in quello marittimo e della logistica.

**Keywords.** Comunicazione Quantistica, QKD, fibre ottiche

## 1 Introduzione: La Crittografia Quantistica

La crittografia quantistica rappresenta un cambio significativo nella sicurezza delle comunicazioni, utilizzando la meccanica quantistica per generare e distribuire chiavi di crittografia. A differenza della crittografia asimmetrica tradizionale, vulnerabile agli attacchi dei computer quantistici, la crittografia quantistica garantisce che qualsiasi tentativo di intercettazione alteri gli stati quantistici, rendendolo rilevabile. Questa caratteristica di sicurezza intrinseca è incarnata nella Distribuzione Quantistica delle Chiavi (QKD), che è stata implementata con successo in varie forme in tutto il mondo.

La QKD implica la generazione e la distribuzione delle chiavi utilizzando fotoni, le particelle fondamentali della luce. In virtù dei principi della meccanica quantistica, qualsiasi tentativo di intercettare questi fotoni ne altera lo stato, lasciando tracce rilevabili. Confrontando porzioni della chiave scambiata, il mittente e il destinatario possono rilevare tentativi di intercettazione. Se vengono rivelate discrepanze, indicando una possibile clonazione della chiave, la comunicazione viene terminata. Questo metodo offre un mezzo intrinsecamente sicuro di distribuzione delle chiavi, una pietra miliare della crittografia quantistica.

## 2 Comunicazione Quantistica in Friuli Venezia Giulia: "Quantum FVG" e "QuFree"

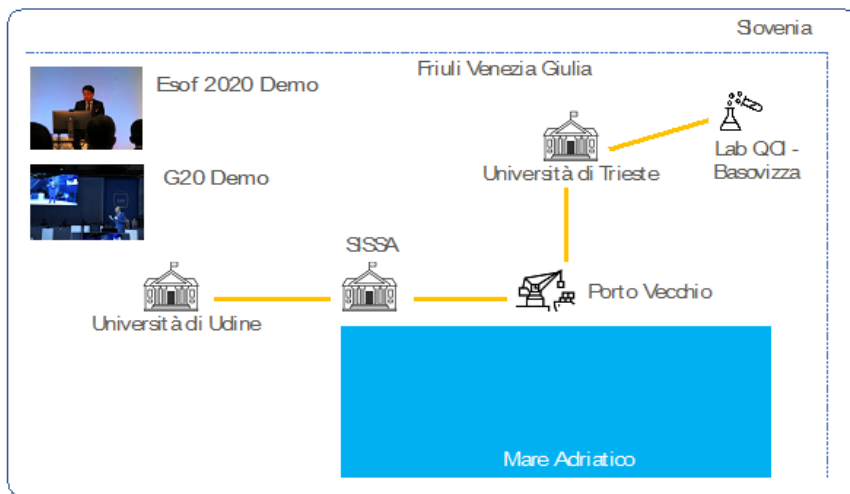
### 2.1 Il Progetto "Quantum FVG"

Il progetto "Quantum FVG"<sup>1</sup>, finanziato dalla Regione Autonoma Friuli Venezia Giulia, ha istituito un laboratorio di comunicazione e informazione quantistica (QCI Lab)<sup>2</sup> presso la sede del CNR-INO<sup>3</sup> a Trieste. Diretto dal Dr. Alessandro Zavatta, questo laboratorio è

dedicato alla ricerca di soluzioni innovative di comunicazione quantistica su fibre ottiche e all'istruzione degli studenti nelle tecnologie quantistiche. Il progetto mira inoltre a costruire la prima infrastruttura quantistica regionale, collegando l'Università di Trieste, con la SISSA (circa 10Km) e con l'Università di Udine (circa 100Km) attraverso una rete di fibre ottiche di proprietà della Regione FVG. È prevista la gestione dell'infrastruttura da parte di LightNet<sup>4</sup>, la rete in fibra ottica regionale che collega le università e la maggior parte degli enti di ricerca regionali.

Il progetto "Quantum FVG" si concentra anche sull'engagement pubblico e sulla dimostrazione delle capacità della tecnologia quantistica. Tra i traguardi significativi ci sono la prima dimostrazione pubblica di comunicazione quantistica in Italia durante ESOF2020<sup>5</sup> a cui ha partecipato il Primo Ministro G. Conte, e la prima comunicazione quantistica al mondo tra Italia, Slovenia e Croazia durante il G20<sup>6</sup> sulla Digitalizzazione e la Ricerca nel 2021.

Fig. 1  
Il progetto "Quantum FVG" ha portato alla istituzione del QCI lab presso la sede di Trieste di INO-CNR e prevede la realizzazione di una infrastruttura quantistica che collega i tre Atenei della Regione



Il laboratorio ha una duplice funzione: avanzare nella ricerca sulla comunicazione e informazione quantistica e fornire una piattaforma educativa per studenti universitari e laureati. Questo doppio focus assicura che la regione non solo contribuisca agli attuali progressi nella tecnologia quantistica, ma prepari anche la prossima generazione di scienziati e ingegneri in questo campo all'avanguardia.

## 2.2 Il Progetto "QuFree"

Il progetto "QuFree"<sup>7</sup>, anch'esso finanziato dalla Regione Friuli Venezia Giulia, mira a estendere la ricerca sulla comunicazione quantistica a canali free-space. La comunicazione quantistica nello spazio libero presenta sfide e opportunità uniche. A differenza delle fibre ottiche, che sono limitate dall'infrastruttura fisica, la comunicazione nello spazio libero può coprire distanze vastissime e raggiungere aree dove la posa della fibra è impraticabile. Tuttavia, deve fare i conti con variabili come le condizioni meteorologiche e le ostruzio-

ni fisiche. Il progetto "QuFree" mira a superare queste sfide, rendendo la comunicazione quantistica sicura più versatile e diffusa.

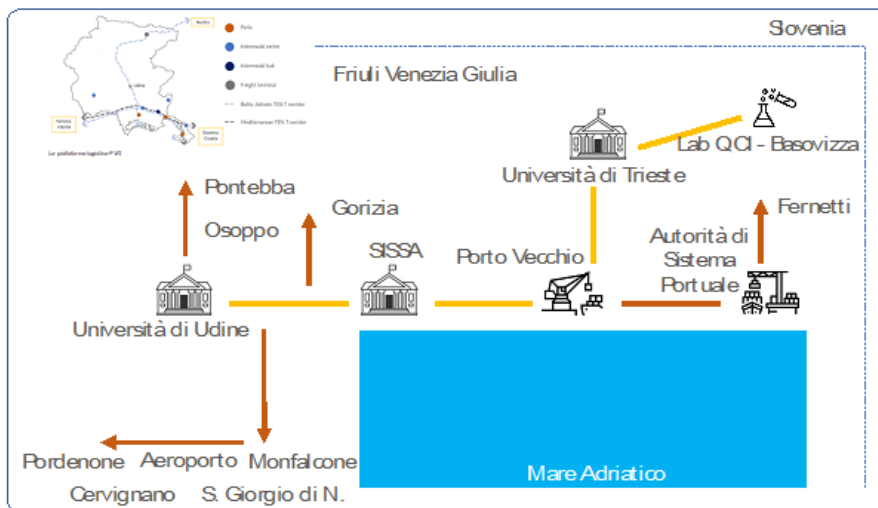
### 3. Prospettive Future

#### 3.1 Integrazione e Collaborazione Avanzata

I piani futuri per i progetti "Quantum FVG" e "QuFree" includono:

- Espandere l'infrastruttura quantistica al sistema regionale della logistica.
- Favorire un ecosistema collaborativo che coinvolga enti pubblici e privati, centri di ricerca e aziende per promuovere e implementare soluzioni di comunicazione quantistica.
- Sostenere la digitalizzazione dei porti moderni e delle navi, migliorando l'efficienza operativa e la sicurezza.
- Trasformare l'infrastruttura quantistica da uno stato pre-industriale a una rete pienamente operativa che soddisfi le esigenze delle telecomunicazioni moderne.
- Collegarsi ad altre infrastrutture quantistiche nazionali ed europee, creando una rete più ampia di canali di comunicazione sicuri.

Fig. 2  
Estensione dell'infrastruttura quantistica prevista dal progetto "Quantum FVG" al sistema regionale della logistica (in fase di studio)



L'espansione dell'infrastruttura quantistica non solo migliorerà le capacità regionali, ma favorirà anche un ecosistema collaborativo che riunisce vari stakeholder.

#### 3.2 Elaborazione di un modello di gestione

Il passaggio dalle attività dimostrative e sperimentali all'utilizzo in pianta stabile della tecnologia di comunicazione quantistica, pone il problema dell'erogazione del servizio, secondo i criteri di affidabilità e resilienza largamente conosciuti dai gestori ed utilizzatori delle infrastrutture classiche. Questi stessi aspetti devono essere indagati ed implementati nelle reti quantistiche, unitamente agli aspetti legati alla presenza di topologie di rete complesse, al loro monitoraggio, alla loro manutenzione, ecc... L'elaborazione di un modello di gestione rappresenta uno degli obiettivi di questi progetti.

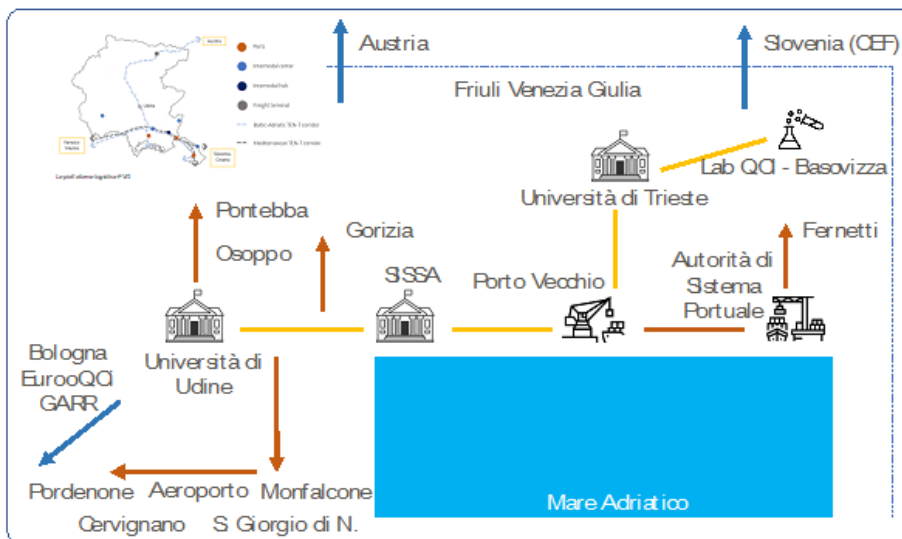
### 3.3 Integrazione delle Reti Nazionali ed Europee

Trieste è destinata a svolgere un ruolo cruciale nelle iniziative nazionali ed europee di comunicazione quantistica:

- Partecipazione al progetto QUID per un'infrastruttura di comunicazione quantistica nazionale.
- Collaborazione nel progetto QUAPITAL per una rete quantistica dell'Europa Centrale.
- Partecipazione ai consorzi guidati da TASI per promuovere le tecnologie quantistiche in Italia.
- Contributo alla Partnership Estesa sulle Scienze e Tecnologie Quantistiche e al Centro Nazionale per il Calcolo ad Alte Prestazioni e il Calcolo Quantistico, entrambi finanziati dal PNRR.

Queste iniziative posizionano Trieste e la regione Friuli Venezia Giulia come attori chiave nel panorama più ampio della comunicazione quantistica europea. Integrandosi con reti nazionali ed europee, la regione può sfruttare una gamma più ampia di risorse ed esperienze, guidando ulteriori progressi nelle tecnologie quantistiche.

Fig. 3  
L'infrastruttura regionale di comunicazione quantistica si inserisce all'interno delle iniziative nazionali (progetto QUID) ed europee (progetto EuroQCI) di comunicazione quantistica



### 3.4 Applicazione nella Logistica Marittima

L'integrazione delle tecnologie di comunicazione quantistica nella logistica marittima è un focus chiave:

- Implementazione di sistemi di comunicazione quantistica sicuri per proteggere contro le minacce informatiche nel settore marittimo.
- Sviluppo di sistemi di navigazione autonoma per navi utilizzando la QKD, con potenziali applicazioni estese ad altri servizi marittimi e terrestri intelligenti.

Il settore marittimo può trarre vantaggi significativi dai sistemi di comunicazione quantistica sicuri. Questi sistemi possono garantire che i dati critici relativi alle operazioni di

spedizione, alla logistica e alla navigazione rimangano sicuri, proteggendosi contro la crescente minaccia degli attacchi informatici.

### 3.5 Potenziamento delle Capacità Educative e di Ricerca

Entrambi i progetti "Quantum FVG" e "QuFree" hanno forti componenti di formazione e ricerca. Istituyendo laboratori all'avanguardia e fornendo accesso ad attrezzature avanzate, questi progetti migliorano la capacità della regione di condurre ricerche all'avanguardia nelle tecnologie quantistiche. Inoltre, offrono preziose opportunità di formazione per gli studenti, contribuendo a coltivare la prossima generazione di scienziati e ingegneri quantistici.

### 3.6 Collaborazione Industriale e Innovazione

I progetti mirano anche a favorire forti collaborazioni con partner industriali, tra cui spiccano Fincantieri, Leonardo, oltre all'Autorità di Sistema Portuale di Trieste. Lavorando a stretto contatto con aziende di vari settori, i progetti "Quantum FVG" e "QuFree" possono stimolare l'innovazione e garantire che le tecnologie sviluppate soddisfino le esigenze del mondo reale. Queste collaborazioni possono portare allo sviluppo di nuovi prodotti e servizi, migliorando ulteriormente la crescita economica della regione e la leadership tecnologica.

## Autori

**Angelo Bassi** [abassi@units.it](mailto:abassi@units.it)

Ha conseguito il dottorato in Fisica all'Università di Trieste nel 2001. Ha lavorato all'ICTP di Trieste, e alla LMU di Monaco. Ora è professore ordinario di Fisica Teorica all'Università di Trieste. Le sue ricerche si concentrano sulla meccanica quantistica. Ha al suo attivo: 150 pubblicazioni su riviste internazionali; relatore su invito in oltre 70 convegni internazionali, organizzarne di 30 eventi scientifici internazionali. Coordina numerosi progetti di ricerca nazionali e internazionali.

**Giorgio Giorgetti** [giorgio.giorgetti@units.it](mailto:giorgio.giorgetti@units.it)

Si è laureato in Ingegneria Elettronica, indirizzo Telecomunicazioni. Dal 2000 lavora presso l'Area dei Servizi ICT dell'Università di Trieste occupandosi di sicurezza informatica e dell'infrastruttura di rete dell'Ateneo. Dal 2008 cura la progettazione e implementazione di LightNet, la rete regionale accademica e della ricerca del Friuli Venezia Giulia di cui è Direttore Tecnico.

## Notes

<sup>1</sup> <https://www.triestequantum.it/quantum-fvg/>

<sup>2</sup> <https://portale.units.it/it/notizie/fisica-quantistica-units-inaugura-due-nuovi-laboratori>

<sup>3</sup> <https://www.triestequantum.it/infrastructure/>

<sup>4</sup> <https://www.units.it/notizie/lightnet/>

<sup>5</sup> <https://www.units.it/news/innovativo-esperimento-quantistico-di-units-ad-esof2020>

<sup>6</sup> <https://onlinelibrary.wiley.com/doi/10.1002/qute.202200061>

<sup>7</sup> <https://www.triestequantum.it/qfree/>

# GPU Offloading in MLOps: Navigating the Multicloud Ecosystem for Flexible AI/ML Deployments

Fulvio Riso<sup>1</sup>, Giuseppe Zangari<sup>2</sup>

<sup>1</sup>Politecnico di Torino, <sup>2</sup>ArubaKube

**Abstract.** The demand for dedicated computational resources for AI tasks is surging. However, the substantial investment required for high-performance GPUs, coupled with increasing market scarcity and extended lead times, poses significant challenges. Simultaneously, underutilization of these resources is a frequent issue. This paper proposes a solution leveraging cloud-native approaches to utilize unused resources available elsewhere, imposing minimal overhead on the originating cluster. We will present how an "origin" cluster can seamlessly peer with a "donating" cluster, offloading tasks to remote hardware with complete transparency and superior efficiency compared to current technologies.

**Keywords.** GPU, Liqo, AI, Efficiency, Offloading

## Introduction

In the ever-evolving landscape of artificial intelligence (AI), the demand for computational resources continues to surge, particularly with the advent of generative AI models. These models, which have gained considerable traction across various fields, introduce new challenges and requirements for efficient deployment. Understanding the computational needs inherent in AI and generative AI is crucial for optimizing performance and ensuring seamless integration into real-world applications.

At the core of AI development lies model training, a process that consumes significant computational power. However, it is essential to recognize that training models is just one facet of the multifaceted activities involved in deploying and managing machine learning operations (MLOps). Beyond training, tasks such as data preprocessing, model validation, and deployment orchestration demand substantial computational resources and operational expertise. Neglecting these aspects can lead to inefficiencies and hinder the scalability of AI systems.

Moreover, the emergence of generative AI, which enables machines to produce novel content such as images, text, and music, presents unique computational challenges. Generative models, such as GANs (Generative Adversarial Networks) and VAEs (Variational Autoencoders), require extensive computational resources during both training and inference stages. The complexity of these models, coupled with the need for large datasets and sophisticated optimization techniques, underscores the importance of robust computational infrastructure.

Managing computational demands in inference tasks can be addressed through various strategies, including optimizing model architectures, implementing hardware acceleration techniques, and utilizing cloud-based solutions for scalable computing. However, the initial investment required for high-performance GPUs or dedicated hardware can be substantial, representing a significant upfront cost. This expenditure is often justified by the promise of enhanced computational power and accelerated AI model training.

Despite the allure of high-performance GPUs, underutilization issues frequently arise. GPUs may not operate at full capacity around the clock, leading to financial inefficiencies. Businesses must grapple with the challenge of maximizing GPU utilization to optimize their return on investment. Conversely, peaks in demand can overwhelm current GPU offerings, particularly in on-premises clusters, resulting in substantial backlogs for users such as scientists and students.

The solution proposed in this paper leverages unused resources available elsewhere, with minimal overhead on the originating cluster. This approach aims to address underutilization and demand peaks, thereby enhancing efficiency and scalability in AI operations.

## 1. Dynamic resource onboarding and cloud bursting

Dynamic allocation is a pivotal strategy for optimizing efficiency and cost-effectiveness in resource management. By dynamically allocating GPU resources in response to real-time demand, organizations can maximize utilization and minimize underutilization, thereby enhancing overall performance while controlling expenses.

However, implementing dynamic resource allocation presents its own set of challenges. One significant challenge is ensuring that Kubernetes clusters can be dynamically extended without necessitating a full reinstallation of the existing infrastructure (Marino et al. 2023). Creating a new cluster (e.g., with Terraform) dedicated to current AI/ML processing on different infrastructure (e.g., another cloud provider) to absorb demand peaks is often impractical. This approach requires replicating all applications and data from the originating cluster, leading to wasted time, resources, and increased costs and operational burdens.

A more effective solution involves adopting robust systems and processes capable of seamlessly integrating new, external GPU resources into the existing cluster. This approach allows the current cluster to extend geographically and integrate with other underutilized clusters that have free GPUs available at any given moment. Adding elasticity to the originating cluster enables it to dynamically offload GPU workloads across the infrastructure, scaling resources up or down in response to changing demands by leveraging available free GPU resources elsewhere.

Such a mechanism allows research infrastructures to optimize GPU utilization proactively, resulting in significant cost savings by avoiding unnecessary hardware investments and minimizing operational overhead. By aligning GPU resources with workload requirements in real-time, research institutions can achieve enhanced agility, resilience, and cost efficiency. This positions them as frontrunners in the competitive landscape of GPU-driven environments, and ultimately in innovation.

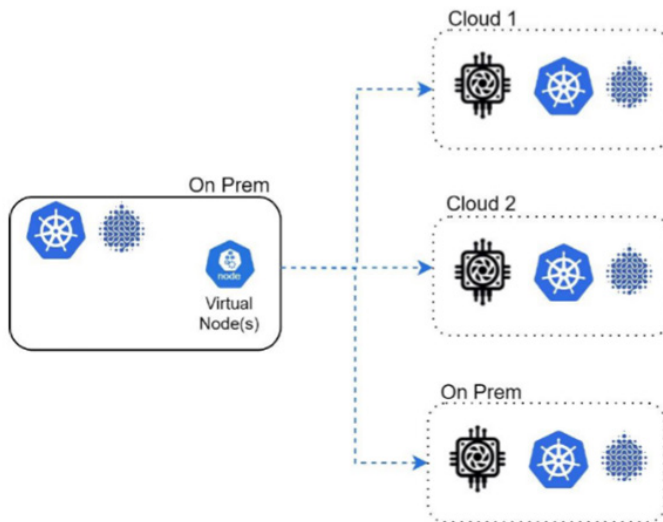
## 2. Kubernetes and Liko

In this context, Liko (Iorio et al. 2023), emerges as a cutting-edge technology that empowers organizations to harness the full potential of their Kubernetes clusters. Designed to facilitate seamless interconnection between Kubernetes clusters, Liko plays a pivotal role in enabling efficient GPU offloading and resource optimization across distributed environments.

At its core, Liko serves as a bridge, connecting Kubernetes clusters regardless of their location or underlying infrastructure. By abstracting cluster boundaries, Liko creates a highly dynamic, unified, federated environment where resources, including GPUs, can be shared and utilized dynamically.

Liko benefits from sophisticated orchestration capabilities that enable the efficient distribution of container workloads across disparate clusters. Leveraging advanced scheduling algorithms and network overlays, Liko orchestrates the deployment and execution of GPU-accelerated applications, ensuring optimal resource utilization and performance. One of Liko's key strengths is its ability to operate seamlessly in multicloud environments, where Kubernetes clusters span multiple cloud providers as well as on-premises clusters. By abstracting cloud-specific complexities and providing a unified management interface, Liko empowers organizations to leverage GPU resources across diverse cloud environments, driving agility and flexibility in their operations.

Fig. 1  
An On Prem cluster is offloading AI tasks to GPU-equipped clusters located in different environments



## 3. An Example of Full Stack Solution

A comprehensive solution to the above problem is emerging that amalgamates Kubernetes, Liko, Jupyter Notebooks, and MetaFlow<sup>1</sup>. Each component brings unique capabilities to the architecture, collectively forming a potent framework for advanced data science and AI endeavours.

Kubernetes serves as the cornerstone of this solution, offering powerful orchestration capabilities for containerized workloads. With features like automated scaling, self-hea-

ling, and resource management, Kubernetes streamlines deployment and management processes, ensuring optimal performance and efficiency across diverse computing environments. Ligo complements this by providing seamless interconnection between Kubernetes clusters, facilitating the creation of easily managed, highly dynamic federated computing infrastructures.

Jupyter Notebooks, renowned for their interactive and collaborative nature, provide a versatile platform for data exploration, visualization, and analysis. Supporting various programming languages and libraries, Jupyter Notebooks empower researchers to experiment, iterate, and share insights in real-time, fostering a culture of collaboration and innovation.

MetaFlow, on the other hand, offers a sophisticated workflow management system tailored for machine learning pipelines. With features like versioning, dependency management, and experiment tracking, MetaFlow simplifies the development and deployment of machine learning models, accelerating time-to-insight and facilitating reproducibility. At the forefront of this transformative initiative is a prototypical implementation currently underway at Politecnico di Torino. Leveraging the combined capabilities of Kubernetes, Ligo, Jupyter Notebooks, and MetaFlow, researchers at Politecnico di Torino are exploring new frontiers in collaborative research and computational science.

However, the potential of this solution extends beyond individual institutions. By embracing a fully federated infrastructure that includes HPC main players and aggregates resources across a wider (e.g., national) research community, additional benefits can be realized. A federated approach enables the pooling of resources, facilitates cross-institutional collaboration, and maximizes the collective impact of research efforts, ultimately driving innovation and advancing knowledge across diverse domains.

## 4. Conclusions

The advancement of AI and ML technologies necessitates efficient and scalable computational resources, with GPUs playing a pivotal role. However, fluctuating GPU utilization presents challenges, including high ownership costs and resource underutilization. Ligo offers a solution by enabling dynamic GPU offloading in a multi-cloud environment, optimizing resource use and reducing operational expenses. By leveraging remote Kubernetes clusters, Ligo facilitates seamless and efficient ML deployments tailored to fluctuating computational demands, mirroring the “resource sharing” model popularized by services like Airbnb.

This paper addresses the operational inefficiencies in traditional GPU management within MLOps, demonstrating how Ligo provides a scalable, cost-effective, and low-code solution. It highlights the significant business impact of adopting a multi-cloud approach, emphasizing enhanced flexibility and the ability to adapt swiftly to research needs. This exploration showcases the competitive advantages offered by such technological agility, including cost reductions, operational efficiency, and high responsiveness to the needs of the academic community.

## Bibliographic References

Iorio M., Riso F., Palesandro A, Camiciotti L., Manzalini A. (2023) Computing without borders: The Way Thowards Liquid Computing, IEEE Transaction on Cloud Computing (vol. 11, no. 3), pp 2820-2838

Marino J., Camiciotti L., Cheinasso F., Olivero A., Riso F. (2023), Enabling Compute and Data Sovereignty with Infrastructure-Level Data Spaces, ESAAM '23: Proceedings of the 3rd Eclipse Security, AI, Architecture and Modelling Conference on Cloud to Edge Continuum (October 2023), pp 77-85

## Authors



Fulvio Riso [fulvio.riso@polito.it](mailto:fulvio.riso@polito.it)

Fulvio Riso is full professor at Politecnico di Torino. Born in Saluzzo, Italy on November 15, 1971, he shares his birthday with the announcement of the Intel 4004 chip. Fulvio completed his BSc in Computer Engineering from Politecnico di Torino in July 1995 and got his PhD in Computer Engineering from the same institution in January 2000. His academic journey has been marked by significant contributions in the field of cloud computing, edge computing, network functions virtualization, and software-defined networking. He greatly contributed to open-source software, starting many successful project such as WinPcap, the de-facto packet capture library for Windows, and many others. He recently started the ArubaKube spin-off of Politecnico di Torino, where he serves as Chief Innovation Officer.

Giuseppe Zangari [giuseppe.zangari@arubakube.cloud](mailto:giuseppe.zangari@arubakube.cloud)

Giuseppe Zangari (born in 1982) graduated from the Politecnico di Torino and holds an EMBA from the Graduate School of Management at Politecnico di Milano. He has held various leadership positions in global software organizations like Nokia and Pirelli, in Italian SMEs and in Politecnico di Torino, leading the development of business effective solutions with technologies ranging from IoT to cloud computing and AI. He is an expert of digital transformation, a startup mentor, and he also served as Innovation Lead. At ArubaKube, he is responsible for maximizing the software project's value, serving concurrently as Product and Business Development Lead.



## Notes

<sup>1</sup>The described tools represent a possible example of deployment, with several alternatives (e.g., KubeFlow vs. MetaFlow) available to adapt to diverse application and usage scenarios.

# Strategie per la progettazione della ricerca finalizzate alla fruizione e valorizzazione dei beni culturali

Sara Alimenti, Elena Gentilini

Università degli Studi di Perugia CeDiPa

**Abstract.** Il 'Centro di ricerca per innovazione, digitalizzazione, valorizzazione e fruizione del patrimonio culturale e ambientale' (CeDiPa), finanziato dal Fondo complementare PNRR per le aree del terremoto del 2009 e 2016 si pone come hub di sperimentazione per rispondere alla grande frammentazione delle iniziative sia volte alla conservazione e alla tutela del patrimonio, che dedicate alla sua valorizzazione e fruizione. In questo contesto, l'Ufficio Progettazione della Ricerca lavora nella problematizzazione ed elaborazione di metodologie e percorsi a supporto della ricerca e del suo dialogo con la società specifici per il settore dei beni culturali e ambientali

**Keywords.** Digitalisation cultural heritage, environmental heritage, research management, digital infrastructures

## Introduzione

La trasformazione digitale ha determinato un rinnovamento profondo di contenuti e metodologie nella ricerca sul patrimonio culturale e ambientale, sollecitando l'apertura di linee di indagine innovative nate dalla contaminazione fra saperi orientati agli aspetti informatico-tecnologici e saperi concentrati sulla dimensione storico, artistica e antropologico-culturale del patrimonio. L'estensione e la stessa natura dei problemi legati allo sviluppo della digitalizzazione offre opportunità rilevanti in termini conoscitivi su questioni inedite legate alla neo-materialità, alla non-materialità e alla cristallizzazione del patrimonio digitalizzato. Dal canto loro, anche le istituzioni e le organizzazioni titolari del patrimonio stanno riorientando gli indirizzi strategici per tentare di rispondere alle nuove esigenze, costruendo sinergie tra loro e con le imprese dei settori culturale e creativo, anche per rispondere alla grande frammentazione sia delle iniziative volte alla conservazione e alla tutela del patrimonio, sia dedicate alla sua valorizzazione e fruizione.

## 1. Il Centro di ricerca per l'innovazione, digitalizzazione, valorizzazione e fruizione del patrimonio culturale e ambientale' (CeDiPa)

Il nuovo Centro di ricerca per innovazione, digitalizzazione, valorizzazione e fruizione del patrimonio culturale e ambientale' (CeDiPa) è un centro interuniversitario nato nel gennaio del 2023 su iniziativa dell'Università degli Studi di Perugia in partnership con l'Università degli Studi dell'Aquila, l'Università degli Studi di Teramo e l'Università Politecnica delle Marche. Finanziato dal 'Piano Nazionale di Ripresa e Resilienza – Fondo complementare, Programma unitario di intervento per le aree del terremoto del 2009 e

2016, Misura B, Sub-misura B.4 Centri di ricerca per l'innovazione', intende rispondere a tutto tondo ai bisogni emergenti legati al processo di trasformazione digitale del patrimonio. Il CeDiPa, che diventerà pienamente operativo nel corso del prossimo triennio, sarà dotato di una infrastruttura tecnologica all'avanguardia ha fin dall'inizio una prospettiva interdisciplinare, tenendo insieme i domini tecnologici con le scienze umane e sociali, per affrontare la complessità dei temi legati alla trasformazione digitale del patrimonio in tutte le sue molteplici declinazioni. Accanto a questo, il CeDiPa, cerca di interpretare al meglio il ruolo assegnato oggi alle istituzioni scientifiche europee, per coniugare nel proprio mandato le attività di ricerca con le iniziative volte a generare impatti significativi anche al di fuori del mondo accademico, in ambito sociale, economico e ambientale. Per questo il CeDiPa è un Centro aperto al territorio e al dialogo con tutte le sue diverse componenti. Il CeDiPa è organizzato in 5 sezioni, dedicate rispettivamente a:

- Patrimonio Architettonico e Urbanistico
- Patrimonio Archeologico Storico e Artistico
- Patrimonio Culturale Immateriale
- Patrimonio Ambientale e Biodiversità
- Tecnologie per la fruibilità e valorizzazione del patrimonio culturale e ambientale.

## **2. La specificità dell'esperienza dell'attività di progettazione della ricerca nell'ambito del Cedipa**

La definizione delle strategie del CeDiPa è affidata ad un Comitato Tecnico Scientifico, formato da professori e ricercatori dei quattro atenei partner e sviluppata poi dai ricercatori in sinergia con l'Ufficio Progettazione della Ricerca del CeDiPa. Obiettivo generale dell'Ufficio Progettazione della Ricerca (UPR) è quello di supportare, con un' impostazione di Learning Organisation (Örtenblad, 2020), i ricercatori nel tentativo di affrontare le sfide attuali relative alla ricerca nel settore del patrimonio culturale e ambientale in una prospettiva interdisciplinare, non rinunciando tuttavia alle specificità degli ambiti disciplinari specifici, considerando cioè entrambe le dimensioni del progresso scientifico essenziali per sviluppare conoscenze e competenze-chiave per offrire soluzioni a problemi tecnologici e sociali complessi.

In particolare, l'UPR sviluppa competenze e strumenti per:

- supportare l'attività dei ricercatori del CeDiPa nelle fasi di pre-award e post-award;
- diventare punto di riferimento a livello regionale, nazionale ed europeo su temi e contenuti che riguardano la trasformazione digitale del patrimonio culturale e ambientale;
- stimolare la definizione dei possibili percorsi di impatto della ricerca del CeDiPa, interrogandosi su come facilitare, sviluppare, indurre, misurare il rapporto fra ricerca e società (Reeta et al., 2019);
- contribuire all'aumento di fiducia verso la scienza e le sue acquisizioni, favorendo una comunicazione reciproca tra scienza e società;
- sviluppare percorsi di ricerca transdisciplinari e trans-settoriali orientati sia alla ricerca che all'impatto, definendo criteri e strumenti che escludano la competizione con altri percorsi di ricerca (per esempio la ricerca di base), ma favoriscano la complementarie-

tà, superando la fase di “embedness” delle Social Sciences and Humanities nei progetti di ricerca (EASSH, 2024) attraverso un costante affinamento della capacità di porre i problemi di ricerca in modo interdisciplinare e di creare forme di collaborazione a lungo termine e sistemiche sia con specifiche organizzazioni che con i cittadini in senso lato.

### **3. Obiettivi e valore aggiunto dei progetti attualmente 'in corso' per il settore dei beni culturali e ambientali**

Il panorama dei progetti su cui il Centro lavora evolve attraverso un lavoro costante di raccolta dati e orientamento strategico, ma gli esiti delle attuali attività progettuali, di seguito elencati, fanno già emergere alcune potenzialità di sviluppo orientate a:

- Digitalizzazione e Promozione delle Opere Salvate In Territori Interessati da calamità naturali e ruolo dei depositi:

- creare un sistema integrato tra Università, Sovrintendenze ed enti pubblici e privati coinvolti nel recupero e valorizzazione in quattro regioni colpite dai sismi (Umbria, Marche, Abruzzo, Lazio) a partire dai depositi;
- definire criteri di documentazione, catalogazione, archiviazione, fruizione e conseguente promozione del patrimonio citato condivisi tra i depositi e gli Enti responsabili, con indicazioni per la replicabilità;
- sperimentare forme di partecipazione in questo processo;
- creare modelli per la preparazione e gestione del patrimonio in caso di disastro a livello europeo.

- La digitalizzazione e la valorizzazione degli archivi storici tra usi scientifici e risorsa per la comunità:

- integrare percorsi di impatto per la valorizzazione di patrimoni diversificati;
- approfondire il rapporto della Fondazione con il mondo della ricerca e con altri settori, i relativi pubblici attuali e potenziali, e il ruolo nell’instaurare nuove connessioni;
- tenere in considerazione la rapidità delle innovazioni del settore, gli standard nazionali ed europei e le opportunità, prevedendo i passi necessari affinché gli investimenti siano sostenibili nel tempo e adattabili a condizioni in continua evoluzione.

- I processi di digitalizzazione per il patrimonio culturale immateriale nelle aree marginali da una prospettiva antropologica:

- creare un terreno di confronto sul concetto di patrimonio e su come questo sia influenzato dal processo di digitalizzazione, anche in chiave propositiva, attraverso strumenti di co-creazione e collaborazione online e offline;
- affrontare le premesse teoriche ed epistemologiche e le metodologie di interesse per la riflessione sul patrimonio immateriale;
- sviluppare le interdipendenze disciplinari e allargare perciò il Team di ricerca;

- entrare da un contesto nazionale ad uno europeo.
- Digitalizzare il patrimonio ambientale come possibile risposta alla sfida di conservazione e valorizzazione della biodiversità:
- a partire da un tema, quello della biodiversità, centrale nelle politiche e strategie europee e asset del sistema della ricerca, e non solo, particolarmente importante in ambito italiano, contribuire al ricongiungimento di patrimonio culturale e ambientale (Boer, 2020), uno degli obiettivi generali del CeDiPa (già nell'impostazione interdisciplinare delle sezioni e ad esempio nell'apertura a ricercatori con background geografico),
  - mappare il complesso panorama europeo in tema di digitalizzazione del patrimonio ambientale rivolto alla biodiversità ed enucleare i possibili percorsi di impatto (ad esempio in ambito Clima o Energia) e le ricerche di frontiera, valorizzando le specializzazioni ed esperienze del gruppo di ricerca – già interdisciplinare e afferente a diversi gruppi di ricerca.
  - Iniziare un'analisi dei bisogni a livello europeo.

## Conclusioni

All'interno della fase di start up del CeDiPa, l'Ufficio Progettazione della Ricerca e i primi nuclei progettuali di attività si configurano come laboratori per individuare e definire obiettivi, metodologie oltre a raccogliere bisogni, specifici del supporto alla ricerca nel settore dei beni culturali e ambientali orientate al raffinamento delle strategie di ricerca e, direttamente e indirettamente, a problematizzare e approfondire il concetto stesso di patrimonio.

## Bibliografia

Örtenblad, Anders R. (2020). *The Oxford Handbook of the Learning Organization*, Oxford University Press, New York.

EASSH (2024), *Monitoring SSH integration still matters*, Position Paper <https://eassh.eu/Position-Papers/Monitoring-SSH-integration-still-matters~p1376>.

Boer Ben (2020), *The Environment and Cultural Heritage'* in Francesco Francioni, and Ana Filipa Vrdoljak (eds) (2020), *The Oxford Handbook of International Cultural Heritage Law*, Oxford Handbooks, Oxford.

Reetta Muhonen, Paul Benneworth and Julia Olmos-Penuela (2019), *From productive interactions to impact pathways: Understanding the key dimensions in developing SSH research societal impact*, *Research Evaluation*, 1–14, Oxford University Press, Oxford.

## Autori



Sara Alimenti, dottore di ricerca in Storia del pensiero politico europeo, è Research Manager presso il Centro di ricerca per innovazione, digitalizzazione, valorizzazione e fruizione del patrimonio culturale e ambientale (CeDiPa) dell'Università degli Studi di Perugia. Accanto ai principali interessi – il dibattito politico sulla democrazia 'dei moderni' e la storia ambientale, indagata attraverso la prospettiva della storia delle idee politiche e sociali – le sue attività di ricerca nell'ultimo triennio si sono orientate verso le questioni relative alla tra-

sformazione digitale del patrimonio culturale e ambientale e sui riflessi di tali trasformazioni per il 'mestiere di storico'.



**Elena Gentilini.** Laurea in Filosofia e Dottorato di Ricerca in Qualità Ambientale e Sviluppo Economico Regionale presso l'Università degli Studi di Bologna, Dipartimento di Economia, lavora come consulente/formatore per enti pubblici e privati, tra cui Camere di Commercio, Associazioni di Comuni e Università nella progettazione strategica con particolare riferimento a transizione circolare e digitale e a competenze ed ecosistemi per l'innovazione. Gli interessi di ricerca riguardano la governance e le metodologie a supporto dei processi di innovazione, in particolare con approcci sistemici. E' Project Manager del CeDiPa e insegna Project Design Lab presso il Master in Progettazione e accesso ai fondi europei per la cultura, la creatività e il multimediale dell'Università degli Studi di Perugia.

## Un cruscotto di monitoraggio della ricerca basato su dati aperti per l'Università degli Studi di Milano

Dario Basset (ORCID 0009-0005-2320-0341), Stefano Bolelli Gallevi (0000-0002-9680-0461)

Direzione Performance Assicurazione Qualità Valutazione e Politiche di Open Science Università degli studi di Milano (<https://ror.org/00wjc7c48>)

**Abstract.** Progetto di un cruscotto di monitoraggio pubblicamente accessibile, che presenti in forma grafica una serie di dimensioni selezionate relative alle attività di ricerca. Nella prima fase del progetto, il focus sarà sulle collaborazioni dei ricercatori dell'Università viste attraverso le loro pubblicazioni. Le fonti di dati utilizzate nel cruscotto sono due:

AIR - <https://air.unimi.it/> - Archivio Istituzionale della Ricerca, che è l'archivio di dati sulla produzione scientifica dell'Università degli Studi di Milano e costituisce di fatto l'Anagrafe della ricerca dell'Ateneo.

OpenAlex - <https://openalex.org/> - che è il catalogo di pubblicazioni scientifiche aperto e gratuito a maggior copertura, in grado di mettere a disposizione i metadati di più di 250 milioni di pubblicazioni.

Il progetto, sviluppato con la collaborazione attiva del dipartimento di informatica dell'Università degli Studi di Milano, è stato avviato ad aprile 2024 e la fase 1 sarà completata a settembre 2024

**Keywords.** OpenAlex, Open Science, Pubblicazioni, Collaborazioni.

Come altre istituzioni soprattutto internazionali, l'Università degli Studi di Milano sta esplorando la possibilità di monitorare la propria attività di ricerca basandosi esclusivamente su dati aperti. Questo appare necessario a fronte di sistemi proprietari costosi e fortemente sbilanciati verso certe aree geografiche, linguistiche e verso specifiche tipologie di pubblicazione.

La Direzione Performance Assicurazione Qualità Valutazione e Politiche di Open Science dell'Università degli studi di Milano ha quindi fatto partire un progetto per la realizzazione di un cruscotto pubblicamente accessibile che presenti in forma grafica una serie di dimensioni relative alle attività di ricerca. In particolare nella prima release del cruscotto si presenteranno le informazioni sulle collaborazioni dei ricercatori e professori dell'Università attraverso le loro pubblicazioni.

Il cruscotto vuole superare la prassi di monitoraggio delle attività di ricerca basata su database proprietari e a pagamento e poco personalizzabili, i cosiddetti Scientific Knowledge Graph, quali Scopus (<https://www.scopus.com>) e Web of Science (<https://www.we>

bofsience.com). Infatti, oltre al pagamento di una licenza d'uso l'utilizzo di tali strumenti presenta l'ulteriore limite di non poter diffondere i dati alla base delle analisi effettuate e dei risultati ottenuti, rendendo così impossibile un efficace monitoraggio comparativo.

La scelta di utilizzare al posto di questi servizi OpenAlex offre notevoli vantaggi in termini di inclusività, accessibilità, disponibilità e riusabilità dei dati. La copertura è circa doppia rispetto a WOS e Scopus e comprende un maggior numero di opere in lingue diverse dall'inglese e del Sud globale. Anche per queste ragioni il ranking di Leiden lo ha utilizzato come fonte per l'edizione del 2024 (questa edizione ha ancora, parallelamente, la versione basata su dati proprietari, ma a partire dal 2025 ci sarà un passaggio alla sola versione basata su dati aperti).

OpenAlex offre un servizio veloce, moderno e ben documentato, ma soprattutto open, infatti OpenAlex è licenziato con licenza CC0, che consente il più ampio riutilizzo. Ispirato all'antica Biblioteca di Alessandria, OpenAlex è un indice di centinaia di milioni di entità interconnesse in un sistema di ricerca globale.

Con il cruscotto, quindi, si vuole ottenere un monitoraggio delle attività della ricerca della nostra istituzione totalmente trasparente e basato su strumenti del tutto aperti. In questa maniera, non ci saranno spese aggiuntive oltre lo sviluppo del cruscotto stesso, e i risultati del monitoraggio possono essere facilmente condivisi con tutta la comunità scientifica e chiunque sia interessato. Il monitoraggio incrocia OpenAlex (<https://openalex.org/>) e AIR (<https://air.unimi.it/>) come fonte dati e per la parte di rappresentazione grafica utilizza strumenti open source. La scelta è quindi aperta by default, come raccomandato dalla dichiarazione di Barcellona (<https://barcelona-declaration.org/>) di cui l'Università di Milano è uno dei primi firmatari in Italia

Il repository istituzionale AIR (<https://air.unimi.it/>) è la fonte dei dati anagrafici dei docenti e ricercatori in attività dell'Ateneo: per ciascun autore viene cercato in OpenAlex il corrispettivo utilizzando l'identificativo ORCID, e vengono poi scaricate tutte le pubblicazioni associate dell'ultimo quinquennio, dalle quali vengono poi ricavate le collaborazioni utilizzando i coautori e le loro affiliazioni; infine il risultato viene esposto online.

## I dati in OpenAlex

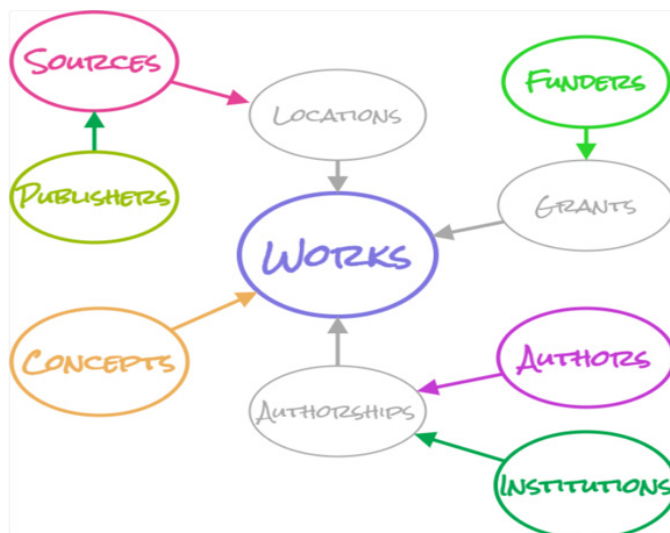


Fig. 1  
<https://docs.openalex.org/api-entities/entities-overview>

La figura mostra i dati disponibili in OpenAlex.

Le entità in gioco sono di seguito descritte:

- Authors sono le persone che creano pubblicazioni (works)
- Works articoli, libri, dataset, tesi
- Funders Organizzazioni che prestano fondi alla ricerca
- Publishers Compagnie ed organizzazioni che distribuiscono le pubblicazioni
- Concepts Gli argomenti delle pubblicazioni
- Institutions Le istituzioni, cioè Università, enti di ricerca ed altre istituzioni
- Sources I giornali ed altri database che ospitano le pubblicazioni

## Funzionalità della prima release

Le ricerche che possono essere effettuate sono:

- Collaborazioni (numero di autori, numero di paper) dell'Università degli Studi di Milano con altre istituzioni in tutto il mondo, raggruppate per Paesi
- Collaborazioni (numero di autori, numero di papers) di ricercatori dell'Università di Milano con altre istituzioni di tutto il mondo, raggruppate per Paesi. Scelta interattiva del ricercatore.
- Collaborazioni (numero di autori, numero di papers) di gruppi di istituzioni di tutto il mondo con l'Università degli Studi di Milano. Scelta interattiva delle istituzioni.

Il cruscotto è interattivo e permette di eseguire ricerche scegliendo lo specifico elemento (il ricercatore, il paese) restituendo grafici e dati tabellari di rendicontazione riguardanti le pubblicazioni scientifiche.

La visualizzazione dei risultati si compone di diverse viste. Nella prima vista vengono illustrate geograficamente le collaborazioni degli autori con altri paesi nel mondo.

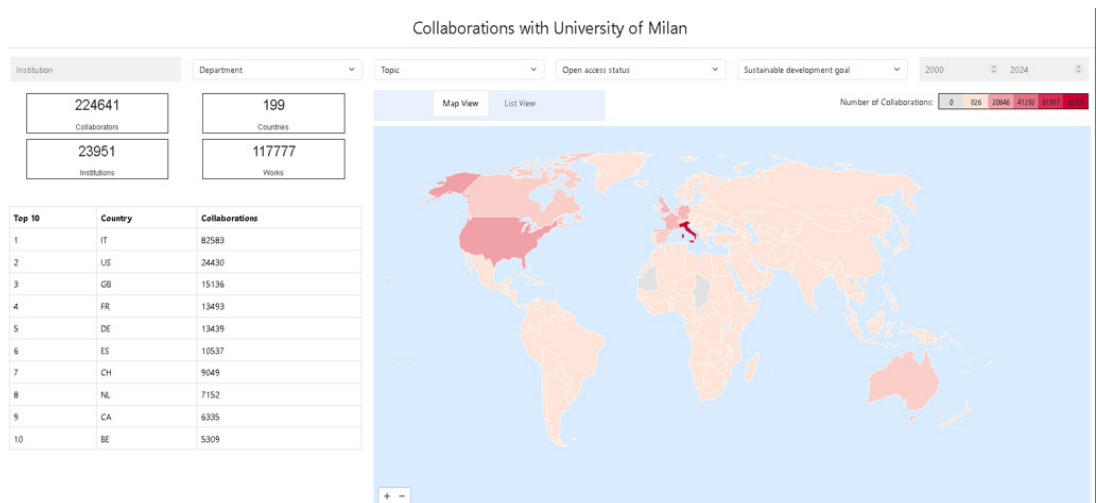
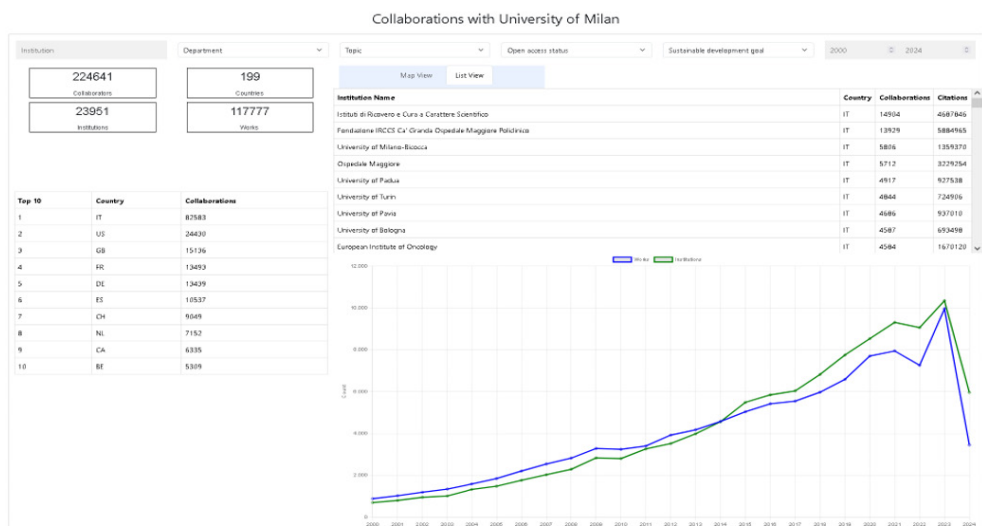


Fig. 2

Vengono mostrati il numero di collaboratori, di istituzioni e di pubblicazioni effettuate in maniera collaborativa dall'Ateneo o dal dipartimento specifico. Esistono inoltre alcuni filtri che hanno effetti selettivi sulla visualizzazione in maniera immediata.



Le stesse informazioni possono essere ricavate dalla vista tabellare come da esempio. Tale vista mostra anche un grafico a linee che descrive l'andamento nel tempo. Anche in questo caso i filtri agiscono in maniera selettiva sui dati, mostrando le viste relative alle scelte effettuate.

## Tecnologie utilizzate

Dal punto di vista tecnologico, si è optato per una scelta completamente open-source, utilizzando librerie e strumenti di sviluppo facilmente disponibili su internet. Il progetto è stato realizzato costruendo un applicativo che usa le api di OpenAlex per ricavare i dati relativi ai ricercatori come censiti in AIR, con i quali alimenta un database da cui un altro applicativo in javascript costruisce l'interfaccia pubblica di esposizione grafica dei dati.

L'applicazione si compone delle seguenti parti:

- Il database utilizzato è postgres 15.
- Il programma che recupera i dati è stato scritto con il linguaggio python versione 3.9 (<https://www.python.org>). Il programma si compone di una serie di moduli che recuperano i dati dapprima dalle tabelle AIR e successivamente da OpenAlex.
- L'interfaccia utente è il cliente ed è stata scritta utilizzando gli strumenti javascript react (<https://react.dev>) e nodeJs (<https://nodejs.org>). In particolare, react realizza l'interfaccia utente attraverso una serie di componenti sul browser, mentre nodeJs agisce da server, ricevendo le chiamate dal client e accedendo al database postgres.

## Autori:

**Dario Basset** presso la Direzione Performance, Assicurazione Qualità, Valutazione e Politiche di Open Science dell'Università degli studi di Milano. Si occupa degli strumenti dell'Ateneo per i dati della ricerca e più in generale dell'open science. È coinvolto nelle attività di formazione che riguardano i dati della ricerca e gli strumenti relativi, e si occupa da più di un anno della qualità dei dati del repository Dataverse.

**Stefano Bolelli Gallevi** lavora presso la Direzione Performance, Assicurazione Qualità, Valutazione e Politiche di Open Science dell'Università degli studi di Milano. Si occupa del repository istituzionale AIR e collabora alle altre attività di Open Science della Direzione.

# Itinerari Digitali

Cecilia Perrone

Istituto centrale per il catalogo e la documentazione - MiC

**Abstract.** Itinerari Digitali è una piattaforma web per la consultazione delle schede catalografiche sul territorio di quattro regioni del sud Italia, individuate nella cornice del progetto PON-FESR 2014/2020. A disposizione degli utenti sono: una mappa interattiva per navigare tra i beni nel loro contesto geografico, e un'applicazione di data entry che permette l'inserimento di nuovi itinerari e la documentazione di punti di interesse non ancora censiti. Per la sua realizzazione è stato prodotto un GEOJson a partire dai dati relativi ai beni immobili e immateriali del Catalogo generale dei beni culturali sul quale si è poi strutturato il database dell'applicativo

**Keywords.** Catalogue, FAIR Data, LOD, Cultural Heritage, UX

## Introduzione

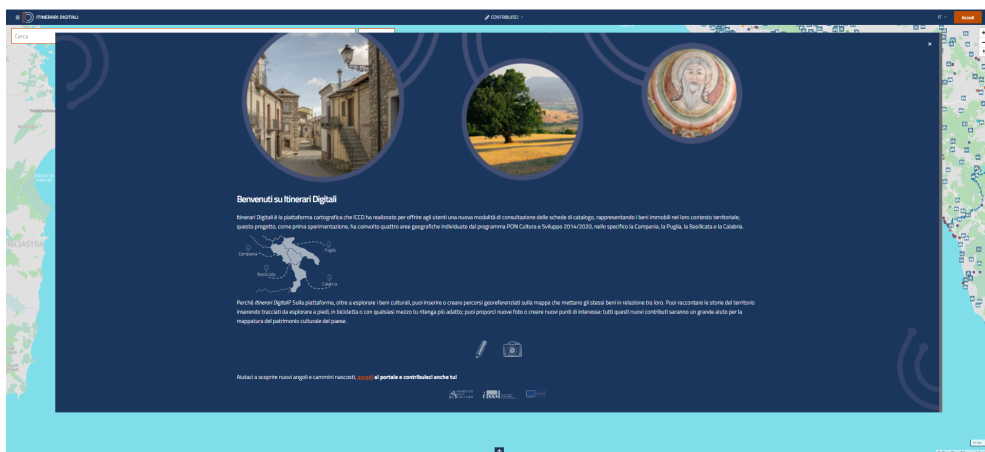
Itinerari Digitali nasce nel quadro delle attività dell'ICCD, finanziate nell'ambito del Fondo Europeo di Sviluppo Regionale attraverso il PON Cultura e Sviluppo 2014/2020. Il progetto proposto dall'Istituto aveva come principale obiettivo la valorizzazione delle attività di catalogazione del patrimonio culturale italiano nel contesto di quattro regioni del meridione (Campania, la Basilicata, la Puglia e la Calabria) territori in cui sono stati individuati particolari punti di interesse definiti Attrattori culturali: siti individuati per la loro rilevanza strategica, secondo precisi criteri di selezione concordati nell' Accordo di Partenariato tra l'Italia e la Commissione europea.

l'ICCD è l'istituto del Ministero della cultura che coordina le funzioni di catalogazione e documentazione del patrimonio culturale nazionale; elabora strumenti e metodologie per valorizzare e rendere accessibile la conoscenza del patrimonio culturale, assicurandone la fruizione attraverso l'utilizzo e la sperimentazione di tecnologie digitali. Gestisce il sistema catalografico nazionale (SIGECweb) e pubblica i risultati delle attività di catalogazione: le schede di catalogo sono navigabili sul portale di consultazione che ne offre anche la versione stampabile (PDF); i dati sono anche esposti in Linked Open Data e l'intero dataset è accessibile dal sito [dati.cultura.gov](http://dati.cultura.gov) dove trovare anche l'endpoint SPARQL per la loro interrogazione (VENINATA, 2020).

Per il raggiungimento dei fini stabiliti nell'ambito sopra descritto, l'ICCD ha realizzato, per la consultazione delle schede di catalogo, uno strumento che parte dalla navigazione del territorio. La totalità dei dati pubblicati, risponde quindi alle più ampie esigenze di consultazione, da quella generalista a quella più specializzata, dimostrando l'intenzione dell'Istituto di distribuire quanto più possibile la conoscenza raccolta e organizzata nel corso dei decenni. L'occasione del progetto europeo offre all' ICCD la possibilità di un ulteriore passo in avanti per ampliare l'accessibilità ai propri dati; viene quindi avviata una profonda riflessione

per lo sviluppo di una applicazione web che consenta la consultazione delle schede su una mappa interattiva, lavorando sui dati geografici che garantiscono il posizionamento del bene catalogato nel suo contesto geografico. Tali dati sono prodotti tramite i servizi di web-GIS sviluppati nel sistema informativo SIGECweb, per consentire la georeferenziazione puntuale, areale o lineare del bene interessato; tramite la georeferenziazione è possibile quindi associare a un dato, secondo uno specifico sistema di riferimento, una coppia di coordinate geografiche (x, y) che individuano il bene sulla superficie terrestre (VENTURA, 2014). Una breve premessa: il presente articolo intende offrire, senza la pretesa di esaurire la trattazione, una panoramica sugli scopi, sul front end e sulle principali soluzioni adottate nell'architettura del software.

Fig. 1  
Itinerari Digitali,  
home page



## I passi del progetto

Come primo passo, porrei un accento sulla scelta del nome del progetto, che richiama il principale scopo del portale: per narrare la storia di un territorio e delle relazioni tra beni o entità culturali radicate in esso, è possibile tracciare dei percorsi fisici sulla mappa, individuando dei punti di interesse o segnando direttamente sulla cartina i percorsi possibili. Gli itinerari prodotti aggiungono valore al patrimonio già documentato dall'attività scientifica di catalogazione, arricchita così da informazioni condivise dagli utenti chiamati a contribuire in prima persona alla valorizzazione dei beni stessi.

Sul portale vengono esposti i risultati di diverse attività, come la perimetrazione e descrizione dei centri e nuclei storici, eseguita in collaborazione con il Dipartimento della protezione civile della Presidenza del Consiglio dei Ministri, attraverso letture comparative di fonti bibliografiche e cartografiche; il lavoro svolto è stato integrato con le fotografie messe a disposizione dall'Aerofototeca nazionale. Parallelamente, è stata avviata una ricerca sulla cartografia storica che rappresentasse il territorio italiano nel corso dei secoli; una particolare attenzione è stata riservata alla collezione David Rumsey (fig.2), che ha costituito un importante spunto di riflessione verso l'utilizzo di mappe storiche a supporto della visualizzazione dei dati sui beni culturali, e che si è rivelata stimolo per future

sperimentazioni.

Ad arricchire ulteriormente l'intero progetto Itinerari Digitali è anche la pubblicazione dei risultati della ricognizione fotografica coordinata dall'ICCD, che ha visto impegnati 16 giovani professionisti nella realizzazione di ben 7000 fotografie inedite che documentano il territorio.

Una ricca e consistente attività ha riguardato poi la documentazione del patrimonio etnoantropologico immateriale, che ha coinvolto quattro catalogatori nella produzione di schede di catalogo e nella realizzazione di risorse documentarie, visive e sonore, valorizzate anche dalla collaborazione con Loquis, una piattaforma di Travel Podcast che ospita i contributi sonori realizzati dagli specialisti.



Fig. 2

fol. 203a, Città di Bari e area circostante, Pirî Reis, d. 1554 (?), Walters Art Museum, W.658, fol 203a, © 2011 Walters Art Museum, used under a Creative Commons Attribution-ShareAlike 3.0 license: <http://creativecommons.org/licenses/by-sa/3.0/>



Fig. 3

Basilica Pontificia Cattedrale di Maria Santissima della Bruna e Sant'Eustachio, Simone Mizzotti, 2022, Basilicata, area territoriale di Matera - Fotografie di Simone Mizzotti, ICCD

## 1. Itinerari digitali, una sfida dei nostri tempi

Tra il 2018 e il 2021 l'ICCD si è impegnato nella pubblicazione in formato linked open data dell'intera banca dati in suo possesso. I dati aperti riguardano oggi circa 3 milioni di sche-

de di catalogo; questa scelta è contestualizzabile nella più ampia adesione ai principi di apertura dei dati prodotti dalle pubbliche amministrazioni, un impegno testimoniato non solamente dalle policy di accesso ai dati e dalle licenze sul riutilizzo degli stessi, ma anche e soprattutto dall'impegno nella produzione di open data secondo i principi cosiddetti FAIR. Infatti, il Catalogo: identifica le proprie entità con URI permanenti che le rendono Findable; è basato su protocolli standard aperti HTTP(S) per l'accessibilità e mette a disposizione un endpoint SPARQL per l'interrogazione del grafo; usa protocolli standard aperti anche per la modellazione della conoscenza (RDF, OWL) per garantire l'interoperabilità; costituisce un nodo del Knowledge Graph pubblicato con licenza CC BY-SA 4.00, licenza che permette ampio riuso dei dati di origine.

L'applicativo Itinerari Digitali beneficia del lavoro svolto sull'intero database sorgente: la pubblicazione del dataset ICCD secondo i paradigmi LOD, ha permesso lo sviluppo di nuove idee per la sua stessa valorizzazione, un'operazione condotta "in casa" ma che dimostra come sia possibile un riuso virtuoso, e privo di particolari ostacoli, di dataset di diversa provenienza, per garantire ai dati stessi una vita lunga e sempre aggiornata (MORO, 2017).

## 2. Il lavoro sui dati e la struttura del sito

La pubblicazione dell'intero dataset del Catalogo ha permesso quindi un'estrazione agevole dei dati di interesse in formato GeoJSON, che è andato a costituire la base di dati del nuovo portale. Il dataset GeoJSON è costantemente allineato al database di origine tramite una pipeline che normalizza e traduce in maniera automatica i dati sorgente; la selezione dei dati si è basata su caratteristiche quali la presenza di dati geografici e la loro presenza nei territori di interesse (le quattro regioni del sud Italia). Tutti i dati provenienti dal Catalogo generale dei beni culturali sono stati rielaborati e riclassificati in dodici macrocategorie, che costituiscono i filtri di ricerca del portale. Il dataset costituito è pubblicato sull'interfaccia ministeriale che espone i LOD interrogabili da qualsiasi applicazione, offerto per usi in contesti analoghi e adatto a ulteriori integrazioni. Il primo applicativo ad aver utilizzato il dataset di Itinerari digitali, espandendo l'interrogazione a tutto il territorio italiano, è stato il Geoportale nazionale per l'Archeologia, realizzato dall' ICA (Istituto centrale per l'archeologia).

L'applicativo è strutturato in pattern a microservizi, una scelta che, in linea con le attuali tendenze nello sviluppo software, lo rende scalabile e garantisce un'esperienza utente continua e solida (CODEMOTION, 2023).

## 3. La mappa

Il portale è stato interamente sviluppato secondo il principio user-centered: la fase preliminare della progettazione ha seguito infatti tutti i passi per garantire una user experience quanto più agile e intelligente possibile (MURPHY, 2018). Sulla base di approfondite riflessioni si è deciso quindi di privilegiare la ricerca che mettesse al centro la navigazione sulla mappa, per permettere all'utente di immergersi nello strumento cartografico esplorando al suo interno le relazioni tra i beni; il visitatore del sito è spinto infatti a spostarsi

sulla mappa e a ricercare i punti di interesse cliccando su di essa e chiedendo all'applicazione di lanciare ricerche nelle vicinanze. L'impostazione della ricerca parte, infatti, dall'indicazione geografica che guida l'individuazione del bene, il cui riferimento potrà essere comunque visualizzato in un convenzionale elenco di record.

La scelta è motivata dallo stesso obiettivo del progetto: fornire al visitatore uno strumento semplice per esplorare il territorio, e costruire su di esso dei percorsi. Per gli scopi di Itinerari digitali è stata appositamente realizzata una mappa Open street Maps (OSM) dedicata esclusivamente alla visualizzazione del patrimonio catalogafico; questo ha permesso l'eliminazione di tutti quei dati non considerati utili agli attuali scopi, come i nomi degli esercizi commerciali della zona. Questa personalizzazione è un sottile dettaglio verso un'esperienza utente sempre più accurata. Per la realizzazione della base cartografica, sono stati importati nell'applicativo dati da sorgenti open source tra cui OSM, OpenAddresses, WhosOnFirst e Geonames; questo permette una ricerca integrata tra il geocoder applicato ai dati di partenza (punti di interesse, schede e itinerari) e gli indirizzi censiti nelle altre banche dati implementate.

Particolarmente accurato è il microservizio GIS, che legge ed elabora i dati del database, restituendoli in un formato vettoriale compresso, permettendo la navigazione veloce nel sito, non appesantito così dalla consistente mole di dati da processare.

#### **4. Crowdsourcing**

Tra i microservizi dell'applicativo, particolarmente prezioso è quello che permette l'autenticazione e gestione degli utenti. L'interazione con l'utenza è forse il punto cardine del progetto; i visitatori del portale, grazie all'implementazione di un altro microservizio dedicato all'upload e al download di risorse multimediali, contribuiscono attivamente alla crescita di Itinerari digitali. È stato previsto un intuitivo strumento di data entry che permette la redazione di itinerari su mappa, corredati da coordinate o da tracciati, di sintetiche schede di punti di interesse e, infine, di itinerari cosiddetti tematici, che permettono di raccontare una storia sui beni senza necessariamente metterli in relazione "fisicamente" sulla mappa. Lo strumento permette poi, a profili di amministrazione, di validare e pubblicare i contributi considerati idonei. La raccolta di informazioni aggiornate e di immagini sul patrimonio culturale, è uno degli obiettivi più preziosi del progetto, che mira ad arrivare a costruire una banca dati quanto più scientificamente attendibile possibile.

#### **5. Conclusioni**

Il portale si configura come una prima sperimentazione di visualizzazione dei dati del Catalogo generale dei beni culturali su una mappa interattiva. È infatti intenzione dell'Istituto centrale per il catalogo e la documentazione, utilizzare il lavoro svolto per lo sviluppo del portale Itinerari Digitali per migliorare il sito di consultazione ufficiale ampliando, sulla base degli assunti qui consolidati, la navigazione del patrimonio culturale a tutto il territorio italiano.

## Riferimenti bibliografici

Moro, L. (2017), Mibact: Così gli Open Data della cultura creano crescita e sviluppo, articolo contenuto in Agenda Digitale.

Veninata, C. (2020) Dal Catalogo generale dei beni culturali al knowledge graph del patrimonio culturale italiano: il progetto ArCo, articolo contenuto in Digitalia 2.

Ventura, S. (2014), Linee guida per il geocoding e la georeferenziazione dei beni in SIGECweb, Istituto centrale per il catalogo e la documentazione, Roma.

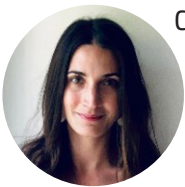
## Bibliografia online

Codemotion, (2023) Design Pattern orientati ai microservizi: cosa sono?, rivista online <https://www.codemotion.com/magazine/it/backend-it/microservizi-cosa-sono/#:~:text=L'architettura%20orientata%20ai%20microservizi,aziendale%20specifico%20tramite%20design%20pattern.> , ultimo accesso giugno 2024.

Murphy, C. (2018), Creating A UX Strategy, Smashing magazine, rivista online, <https://www.smashingmagazine.com/2018/03/creating-a-ux-strategy/> ultimo accesso giugno 2024.

I dati pubblicati rispettano i criteri di visibilità stabiliti in sede di definizione degli standard catalografici; dati particolarmente sensibili vengono infatti filtrati e oscurati alla consultazione pubblica.

## Autori



Cecilia Perrone [cecilia\\_perrone@hotmail.it](mailto:cecilia_perrone@hotmail.it)

Sono laureata in storia dell'arte presso l'Università Sapienza di Roma, collaboro con l'ICCD dal 2020, seguendo le attività di catalogazione sul territorio del sud Italia e supportando i funzionari nelle diverse attività dell'Istituto. Parallelamente all'attività lavorativa, proseguo gli studi con la Scuola di specializzazione in beni archivistici e librari della Sapienza di Roma. Ho avviato da qualche tempo un percorso di approfondimento sulle tecnologie del Semantic Web e della rappresentazione della conoscenza tramite tecnologie informatiche.

# Sovranità digitale: DPO, protezione dei dati e cybersicurezza

Lucio Badiali

INGV Istituto Nazionale di Geofisica e Vulcanologia

**Abstract.** La sovranità digitale ha tra le sue iniziali priorità la difesa del perimetro cibernetico al cui interno si trova anche il dato personale, vero asset strategico di ogni pubblica amministrazione (PA) l'insieme delle quali formano i punti di contatto tra cittadino e Stato. La protezione del dato richiede diverse abilità specialistiche distribuite spesso tra più figure

**Keywords.** RDT, Referente per la cybersicurezza, CIO, DPO, PA

## Introduzione

Il DPO non può ridursi a essere un esperto di sterile normativa, perché la data protection è una prassi, e deve iniziare a palesarsi come una delle voci che analizza il rischio e che collabora con gli altri attori implicati nella difesa degli asset. E' un ruolo sì ausiliario ma che può supportare il passaggio alla tanto attesa resilienza del perimetro cyber della PA.

## 1. Sovranità digitale: DPO, protezione dei dati e cybersicurezza

Il ruolo del DPO in un soggetto pubblico, sia un ente di ricerca nazionale che una infrastruttura di ricerca (IR) europea, è chiaramente simile. Le sfumature emergono nell'interazione con i vari delegati e gli uffici distribuiti nel territorio dell'Unione, e nei paesi che collaborano pur essendone fuori. Nel caso di EPOS l'interazione è con un policy team in cui nell'ultimo periodo, prima delle periodiche rotazioni, il responsabile per le politiche di sicurezza ICT e l'esperto legale erano in UK in quanto tra i partecipanti compare il BGS. Abbiamo, pertanto, vissuto anche noi il pre- e il post- Brexit nella data protection. La sede legale di EPOS è in Italia per cui vale, oltre il fondamentale GDPR, la normativa italiana che va condivisa con i partner stranieri per determinare le best practice in materia.

Nonostante la complessità di una IR, gli asset e i processi sono limitati e ben distinguibili rispetto alla realtà nazionale che deve scontare, almeno nel nostro caso, frequenti riorganizzazioni con spostamenti e cambi di mansione del personale.

L'attività privacy non si esprime solo nella verifica degli agreement e nelle relative informative, nel registro trattamenti, lato più epidermico e visibile, bensì nel lavoro di pianificazione che ha nei due termini la ragione della sua azione: protezione e dati personali.

I dati sono l'asset fondamentale, il vero patrimonio dell'ente, con un valore difficilmente stimabile e che è sottoposto a potenziali aggressioni per essere sottratto (data leak e data breach), modificato o cancellato trovandosi così a non garantirne la celebre triade alla base

della sicurezza delle informazioni: riservatezza, integrità e disponibilità.

Una infrastruttura di rete ed i suoi servizi possono essere messi fuori uso temporaneamente. La protezione del dato è una azione da pianificare su più dimensioni e da realizzare ben prima dell'evento sinistro, infatti il dato una volta esfiltrato è perso per sempre e diventa patrimonio di altri, mentre i servizi di rete alla fine, comunque, tornano su.

Il DPO oggi non è solo l'esperto della normativa di settore, ma deve capire il pensiero ed il modo di operare dell'esperto di sicurezza informatica, deve cioè necessariamente muoversi agevolmente in entrambi i temi e legarli con una approfondita conoscenza dell'analisi del rischio. Quando il dato è compromesso e si attiva un confronto con l'Autorità garante, e non solo quella, non serve parlare di leggi e norme, perché la domanda è solo una e richiama il principio di accountability: "cosa avete fatto prima dell'evento?", "Come avete analizzato lo stato del vostro sistema e messo ragionevolmente in sicurezza gli asset?". In quel "ragionevolmente" si gioca la partita. Un DPO che asseconi le scelte tecniche di sicurezza, perché non in grado di comprenderle e non interagisce in maniera dialettica con gli specialisti, non fornisce alcun valore aggiunto. Né può esercitare il suo ruolo sul tema assicurandosi, passivamente, con i vari specialisti che i dati siano ben protetti. Sarebbe come chiedere all'oste: "come è il vino?". E la sua difesa si fa debole se non inutile.

In tema di sicurezza informatica e cybersicurezza, soprattutto nel nostro Paese, si tende ad allungare a dismisura la catena di comando. In questo momento, vigente la NIS, siamo in attesa dell'attuazione in autunno della direttiva NIS2, ma si è sentito il bisogno di anticipare con un "decreto cybersicurezza" che inserisce una nuova figura quella del referente per la cyber sicurezza, fondamentale elemento per difendere le pubbliche amministrazioni dalle minacce digitali e assicurare la protezione dei dati sensibili, nonché la continuità delle operazioni. Come se prima non ci fosse stata una simile esigenza. Tanto per ricordare, un inciso che vale come esempio, anche con la NIS un settore strategico come la sanità, è stato ritenuto una filiera strategica che rientrava completamente nell'ambito di applicazione della disciplina di cybersicurezza, ma nonostante questo, si sono vissuti drammatici attacchi alle strutture sanitarie con servizi in down completo ed esfiltrazioni di dati personali particolari (e parliamo di casi noti finiti spesso in cronaca, e alcuni sono casi ancora aperti). Se nel recente passato gli effetti sono stati negativamente impattanti, forse il problema potrebbe trovarsi negli investimenti, nelle scelte, o proprio nella catena di controllo e comando. Stratificare e complicare la normativa è una pratica nazionale tipica e, quando non si riesce a gestire la complessità, purtroppo, se ne si aumenta il livello. A causa delle incomprimibilità dei bilanci, in genere le P.P.AA. devono spesso operare senza variazioni, senza maggiori spese, tanto che la figura del referente può essere il responsabile della transizione digitale. Diversi sono gli specialisti con cui il DPO si confronta per questo tema. Il referente cyber (potremmo vederci una sorta di CISO della PA), il cui testo di riferimento è il succitato recente d.d.l. cybersicurezza (insieme alle varie NIS) ha il suo punto di contatto in ACN, e il responsabile della transizione digitale (RTD), il cui riferimento normativo principale è il CAD e che ha il suo contatto in AgID, sono figure che vengono create per facilitare cambiamenti verso un futuro di resilienza, anchee cibernetica. Da non dimenticare che esiste anche il responsabile informatico, il CIO, equivalente

ormai al sergente in trincea che lotta per avere personale e finanziamenti, ma oltre ai vari direttori ha due nuovi ufficiali in capo. Allo stato attuale, la gestione delle strategie e della prassi della sicurezza ICT è assimilata a un vettore a tre figure componenti (RTD, CISO, CIO) ma che per la situazione della PA italiana, si può ridurre a un vettore che perde, per motivi di finanziamento, una ad una le sue dimensioni fino a collapsare sulla stessa unica componente personale. Generando un paradosso in cui chi deve definire le strategie è anche quello che le applica, cosa che genera sottili conflitti di interessi e identifica controllore e controllato. Una catena di comando in linea di principio forse troppo lunga ma che si può ridurre ad una sola persona con cappelli diversi da indossare in contesti diversi. E in un attimo, siamo di nuovo ai tempi in cui esisteva il solo responsabile IT.

Tornando alla situazione internazionale molto più elastica in termini di operatività e dove è anche più facile reperire risorse umane preparate, esiste una difficoltà nello spiegare la versione sartoriale in cui produciamo normative e berretti da indossare secondo chi si incontra, e allunghiamo i riferimenti autoritativi con cui interagire. Come ben evidenziato da quasi tutti i relatori durante la Conferenza del GARR 2024, un problema sentito risiede proprio nel reperire il personale preparato, spesso anche formato da noi, che decida poi di restare nell'ente di appartenenza.

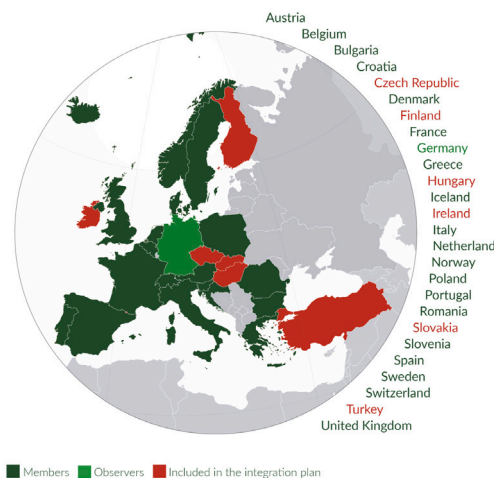


Fig. 1  
EPOS European  
Research Infra-  
structure Con-  
sortium (ERIC)

## 2. Conclusioni

In questo periodo la figura del DPO può rinnovarsi per diventare forse quella che era stata immaginata durante la genesi del GDPR. Grazie alle spinte esterne, dovute pure alla sfida della IA e alle sempre più originali minacce cyber, il suo ruolo si ridefinisce. Per esempio, si nota che a ridosso del decreto cyber è stato accolto un ordine del giorno che impegna il governo a specificare che le pubbliche amministrazioni, almeno quelle centrali, per quanto riguarda la cybersicurezza, debbano coinvolgere il responsabile per la transizione digitale e il responsabile della protezione dei dati. Se si riesce ad usare questa opzione anche noi, se il DPO riesce a parlare la lingua degli esperti in materia cyber senza usare interpreti e traduttori avremo una risorsa in più dove di risorse necessarie non se ne trovano molte. E

questo ha senso perché GDPR e normative cyber hanno uno stesso obiettivo, far adottare misure tecniche e organizzative adeguate, per raggiungere il più elevato livello di sicurezza compatibile con il periodo presente, guardando al futuro.

### **Autore**



Lucio Badiali [lucio.badiali@ingv.it](mailto:lucio.badiali@ingv.it)

DPO di INGV e EPOS.

Laurea in Fisica, specializzazione in cibernetica, Dottorato in Metodologie Fisiche Innovative per la Ricerca Ambientale. Laurea in Scienze politiche, Master Universitario di 2° livello (CMU-2) in Scienze Giuridiche Diritto della Protezione dei Dati Personali.

# NeuroArtP3: un modello di gestione del trattamento dei dati nell'ambito di un Programma di rete dopo l'entrata in vigore del GDPR

Manuela Marengo<sup>1\*</sup>, Irene Sanginisi<sup>1\*</sup>, Matteo Pellini<sup>1</sup>, Andrea Gambino<sup>2</sup>, Luisa Franco<sup>1</sup>, Cristina Grandi<sup>1</sup>, Marco Damonte Prioli<sup>3</sup>, Luca Boni<sup>2</sup>, Antonio Uccelli<sup>1</sup>

<sup>1</sup>Grant and Technology Office, Direzione Scientifica, IRCCS Ospedale Policlinico San Martino, Genova, Italia, <sup>2</sup>U.O.C. Epidemiologia Clinica, IRCCS Ospedale Policlinico San Martino, Genova, Italia, <sup>3</sup>Direzione Generale, IRCCS Ospedale Policlinico San Martino, Genova, Italia

\*i due autori hanno contribuito egualmente alla realizzazione dell'articolo

**Abstract.** L'IRCCS Ospedale Policlinico San Martino di Genova è il coordinatore del Programma di rete ministeriale NET-2018-12366666 "NeuroArtP3", il cui obiettivo principale è lo sviluppo di modelli computazionali di intelligenza artificiale predittivi, preventivi e personalizzati applicati a patologie neuro-degenerative, infiammatorie e tumorali. Date le molteplici esigenze di conformità con i requisiti clinici, etici e privacy legate alla sua realizzazione, il San Martino ha stabilito di coordinare il Programma dedicando specifiche risorse al management scientifico, privacy e amministrativo. Questo approccio si è rivelato un giusto compromesso, consentendo di raggiungere gli obiettivi di ricerca del Programma e al contempo di ottemperare alle prescrizioni etiche e della normativa privacy dopo l'entrata in vigore del GDPR.

**Keywords.** Programma di rete NeuroArtP3, Protocolli multicentrici osservazionali retro/prospettici, GDPR, Software di raccolta dati, Management

## 1. Introduzione

Lo sviluppo di modelli di apprendimento automatico (ML) preventivi, predittivi e personalizzati (P3) attraverso l'analisi dei big data provenienti da vari IRCCS è tra le strategie più attuate nella ricerca sanitaria, per migliorare l'assistenza e la gestione dei pazienti, nonché ottimizzare l'uso delle risorse sanitarie riducendone i costi.

In questo scenario si inserisce il Programma di rete ministeriale NET-2018-12366666 "Intelligenza artificiale dei dati di imaging e clinici neurologici per la medicina predittiva, preventiva e personalizzata (P3) - NeuroArtP3", che mira a sviluppare algoritmi capaci di prevedere gli esiti clinici e la sopravvivenza, personalizzare la terapia e stratificare i pazienti affetti dalle seguenti patologie: Malattia di Alzheimer (AD), Malattia di Parkinson (PD), Sclerosi Laterale Amiotrofica (SLA), Sclerosi Multipla (SM), glioma adulto e pediatrico. Il Programma prevede diversi Working Packages (WP), suddivisi in 5 centri clinici e 2 cen-

tri computazionali (Tab. 1) e si sviluppa attraverso la raccolta e l'analisi ML di big data retro/prospettici, disponibili come da normale pratica clinica.

Tab. 1  
"WP e rispettivi ruoli"

WP e Regioni	Centri	Acronimo	Ruolo	Patologia in studio
WP1 - Liguria	IRCCS Ospedale Policlinico San Martino	HSM	Coordinatore Clinico	AD, PD, SM, SLA
WP2 - Trento	<ul style="list-style-type: none"> <li>• Azienda Provinciale dei Servizi Sanitari</li> <li>• Fondazione Bruno Kessler</li> </ul>	APSS	Clinico	PD, SM
		FBK	Computazionale	Tutte
WP3 - Milano	IRCCS Ospedale San Raffaele	HSR	Clinico	Gliomi dell'adulto
WP4 - Liguria	Istituto Giannina Gaslini	IGG	Clinico	Gliomi pediatrici
WP5 - Firenze	Fondazione Don Carlo Gnocchi	DCG	Clinico	SM
WP6 - Liguria	Dipartimento di Matematica - Università degli Studi di Genova	DIMI	Computazionale	Tutte

In questo articolo si analizzerà il modello di governance adottato da HSM, quale centro coordinatore del Programma, ovvero l'insieme delle scelte, azioni e misure, concordate con i WP, per condurre la ricerca in conformità con i requisiti clinici, etici e privacy.

La realizzazione del NeuroArtP3 si inserisce infatti in uno scenario di significative modifiche normative, a seguito dell'entrata in vigore nel 2018 del Regolamento Generale sulla Protezione dei Dati (GDPR), che ha sostituito il precedente quadro giuridico europeo del 1995, introducendo nuovi obblighi di conformità per lo svolgimento della ricerca scientifica (Mondschein C.F. et al, 2018). All'entrata in vigore del GDPR è seguita la modifica del Codice Privacy (D.Lgs n.196/2003), pertanto il NeuroArtP3 è stato uno dei primi programmi a essere condotto in questo nuovo contesto normativo, affrontando la sfida di bilanciare esigenze scientifiche e privacy.

## 2. Management del NeuroArtP3

Dopo l'approvazione da parte del Ministero e delle Regioni coinvolte, il NeuroArtP3 è stato avviato ufficialmente il 20/08/2020. Fin dall'inizio, vista la sua complessità, HSM ha ritenuto necessario creare un "team di gestione", costituito da due risorse, ognuna per lo specifico ambito di competenza, per coordinarne gli aspetti scientifici, privacy e amministrativi. Questa scelta ha portato a porre l'accento, più che sulle peculiarità dei singoli WP, sulle patologie oggetto di studio del Programma e sulle relative analisi ML. Da questa impostazione, condivisa dai WP, deriva quanto segue:

### 2.1 Protocolli multicentrici osservazionali

Sotto la guida di HSM, per ogni patologia è stato sviluppato un protocollo di studio multicentrico osservazionale, retro/prospettico, con la partecipazione dei WP clinici coinvolti nello studio delle patologie (Tab.1) e dei computazionali. HSM è il promotore di ogni protocollo e ne ha coordinato la stesura a partire da uno stesso template, guidando anche l'armonizzazione della raccolta dati, stabilendo i necessari controlli di qualità e definendo 5 eCRF di patologia, che rappresentano i dati clinici e di imaging disponibili presso i WP. L'arruolamento si articola in due fasi: la prima retrospettiva coinvolge i pazienti con visita iniziale entro dicembre 2019 e costituisce la coorte di test per le analisi ML. La seconda

prospettica coinvolge pazienti con una visita iniziale a partire da agosto 2020 e viene utilizzata per la convalida delle analisi condotte su quella retrospettiva.

Visto che il NeuroArtP3 tratta categorie di dati particolari (art.9 GDPR), come dati clinici, neuropsicologici, biologici, di laboratorio, di imaging, genetici, è stata prestata grande attenzione alla scelta delle misure di sicurezza, stabilendo che la raccolta dati avvenisse mediante i software REDCap (<https://projectredcap.org/>) (Harvey L.A., 2018) per i dati clinici e XNAT (<https://xnat.org/>) per l'imaging, in quanto quest'ultimo permette di caricare, visualizzare e scaricare le immagini raccolte, proteggendo i dati e gestendone l'accesso in conformità alle disposizioni GDPR (Herrick R. et al, 2016). HSM ha installato i software, fornendone l'accesso agli altri WP, secondo specifici ruoli e autorizzazioni (Tab. 2). Inoltre, tutti i dati sono stati raccolti nei software dopo essere stati pseudonimizzati da ciascun WP clinico, che ha assegnato a ogni paziente un codice di identificazione alfanumerico, seguendo criteri stabiliti congiuntamente da tutti i WP. I pazienti arruolati sono stati quindi registrati nell'eCRF mediante il loro codice univoco, così che il WP arruolante fosse l'unico soggetto in grado di identificarli, tramite l'associazione codice-nome (art.4 GDPR). Questa misura ha impedito che i dati fossero direttamente ricondotti ai pazienti, assicurando un congruo livello di tutela privacy.

Tab. 2  
"Profili e autorizzazioni per l'utilizzo di REDCap e XNAT"

	Progettazione	Attivazione utenti	Accesso ai dati	Export dei dati	Controlli di Qualità
<b>HSM</b>	✓	✓	vedere	✓	sviluppo
<b>WP clinici</b>	✗	✗	vedere e modificare	✗	verifica
<b>Centri computazionali</b>	✗	✗	vedere	✓	✗

L'armonizzazione dei protocolli ha portato anche all'allineamento delle informative/consensi informati dei pazienti alla partecipazione allo studio e al relativo trattamento dei dati personali. HSM ha poi coordinato l'armonizzazione di tutti i documenti relativi alla sottomissione ai comitati etici di riferimento.

## 2.2 Adempimenti privacy

Oltre all'impostazione etico/progettuale di cui sopra, HSM ha anche coordinato gli aspetti privacy, programmando e stabilendo i seguenti adempimenti, necessari affinché il Programma fosse GDPR compliant:

- Valutazione dell'Impatto del Trattamento dei Dati (DPIA): ai sensi dell'art.35 del GDPR, prima di avviare qualsiasi tipo di trattamento è stato necessario che i WP clinici, in qualità Titolari Autonomi, redigessero la propria DPIA (Heys M., Smyth R.L., 2020). Per la sua stesura, HSM ha istituito un team ad hoc composto da figure professionali con competenze scientifiche, informatiche e legali, che è rimasto attivo per l'intera durata del Programma, vista la necessità di tenere costantemente aggiornato il documento. Per garantire l'armonizzazione delle sezioni relative alle attività scientifiche promosse nel NeuroArtP3, una volta validata con esito positivo, HSM ha condiviso la propria DPIA con gli altri WP, evidenziando le parti cento-specifiche da adattare alle singole realtà e misure aziendali adottate.

- **Accordo sui Dati Personali e Nomine dei Responsabili del Trattamento:** poiché NeuroArtP3 comporta la raccolta e l'analisi di dati particolari dei pazienti, per massimizzarne il potenziale, garantendone al contempo la conformità al GDPR, i WP clinici hanno deciso di firmare un "Accordo sui Dati Personali", per delineare le reciproche relazioni. Sebbene questo accordo non sia un requisito obbligatorio del GDPR, la sua stipula ha consentito di definire, a livello privacy, i ruoli delle parti e la gestione degli adempimenti da osservare (DPIA, informative e consensi, scelta di misure tecnico-organizzative adeguate, politica di pubblicazione), in un Programma di ricerca complesso come quello di NeuroArtP3. Tutti i WP clinici sono stati designati Titolari Autonomi del trattamento dei propri dati, così da valorizzare le caratteristiche dei centri, garantendone una maggiore autonomia decisionale. Di conseguenza, ciascun WP clinico è il solo responsabile delle proprie attività, condotte nell'ambito del NeuroArtP3. HSM, in qualità di coordinatore del Programma, è qualificato come destinatario dei dati di tutti i WP e ne deve garantirne qualità e sicurezza. I due centri computazionali sono stati invece qualificati come Responsabili del Trattamento dei Dati, ai sensi dell'art. 28 del GDPR e nominati autonomamente dai WP clinici, per condurre le attività ML stabilite nei protocolli di studio. Anche in questo caso, tale adempimento è stato realizzato con il supporto di HSM, che ha predisposto e finalizzato gli atti di designazione dei due Responsabili del Trattamento. Una volta stipulati, gli atti sono stati condivisi come modelli con gli altri WP, in modo che potessero scegliere se adottarli o usare i propri template per nominare i due Responsabili.
- **Accordo di Trasferimento di Materiali (MTA):** in base agli obiettivi specifici del protocollo PD, che prevede l'invio di campioni di sangue da APSS a HSM per l'analisi genetica, quest'ultimo ha redatto un MTA per delineare i termini e le condizioni sotto cui il materiale biologico viene trasferito, inclusi i diritti di proprietà intellettuale, gli usi consentiti, gli obblighi di riservatezza.

### 3. Conclusioni

Un Programma così complesso che, per ogni patologia, vede l'attivo coinvolgimento di diversi WP, sia nella raccolta retro/prospettica di dati tra loro molto eterogenei, sia nell'analisi ML con finalità 3P, ha richiesto un approccio metodologico, regolatorio e privacy estremamente articolato, essenziale per garantire il rispetto dei requisiti propri di tutti i livelli. Per ottimizzare le risorse, HSM ha stabilito di istituire, al suo interno, una figura di coordinamento per la preparazione di tutta la documentazione per sottomettere gli studi ai Comitati Etici coinvolti e per armonizzare la raccolta dati, disegnando le eCRF, i controlli di qualità e pianificando i ruoli dei soggetti coinvolti. Analogamente, ha designato una figura per la gestione dei profili privacy, che si è interfacciata costantemente con il management scientifico al fine di bilanciare i due aspetti, soprattutto a seguito dell'entrata in vigore del GDPR nel 2018.

Questa impostazione ha comportato la necessità di richiedere un'estensione di quasi due anni della durata del Programma (prorogato al 19/04/2025) e una modifica del piano operativo. Tuttavia si ipotizza che il grande sforzo fatto possa proseguire in studi futuri, atti alla validazione clinica dei primi promettenti risultati osservazionali ottenuti dai test sui

dati retrospettivi, che saranno a breve ripetuti sulle coorti prospettiche.

#### 4. Bibliografia

Mondschein CF, Monda C, Kubben P, Dumontier M, Dekker A (2018), The EU's General Data Protection Regulation (GDPR) in a Research Context, *Fundamentals of Clinical Data Science* [Internet]. Cham (CH): Springer; 2019. Chapter 5

Harvey LA (2018), REDCap: web-based software for all types of data storage and collection, *Spinal Cord*. 56(7):625

Herrick R, Horton W, Olsen T, McKay M, Archie KA, Marcus DS (2016), XNAT Central: Open sourcing imaging research data, *Neuroimage* 1;124(Pt B):1093-1096

Heys M, Smyth RL (2020), General data protection regulation: What does this mean for research? *Arch Dis Child Educ Pract Ed*.105(5):296-297

#### Autori



**Manuela Marenco** [manuela.marenco@hsanmartino.it](mailto:manuela.marenco@hsanmartino.it)

Laureata in Biotecnologie, ricercatrice all'Institut Curie di Parigi e nella Medicina Nucleare del San Matteo, dal 2021 è Project Manager Scientifico al San Martino. Il ruolo le ha consentito di approfondire la gestione della ricerca clinica, in particolare le attività di coordinamento delle professionalità coinvolte nella realizzazione di studi. Ha acquisito un'ottima conoscenza degli aspetti di management, normativi e regolatori che governano la ricerca clinica sperimentale e osservazionale.

**Irene Sanginisi** [irene.sanginisi@hsanmartino.it](mailto:irene.sanginisi@hsanmartino.it)

Avvocato, iscritta all'Ordine degli Avvocati di Genova. Ha maturato, tra le altre, esperienza in materia di diritto della privacy e della protezione dei dati personali nonché in materia di corporate compliance. Nel 2021 ha frequentato presso la Luiss Business School il corso avanzato "Governance della Privacy", conseguendo la relativa certificazione. Dal 2023 collabora con il San Martino per la gestione dei profili privacy del Progetto NeuroArtP3.

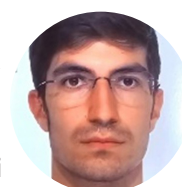


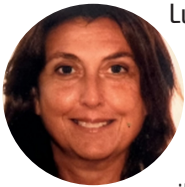
**Matteo Pellini** [matteo.pellini@hsanmartino.it](mailto:matteo.pellini@hsanmartino.it)

Dottorato presso l'Università degli Studi di Genova dal 2020 e abilitato all'esercizio della professione forense, dal 2021 è Privacy Officer presso la Direzione Scientifica del San Martino. Occupandosi principalmente di ricerca, il suo ruolo gli ha dato l'opportunità di approfondire il tema della protezione dei dati personali negli studi clinici e nei progetti nazionali ed europei, nell'ottica di coniugare la ricerca con il rispetto della normativa italiana sulla privacy e del GDPR.

**Andrea Gambino** [andrea.gambino@hsanmartino.it](mailto:andrea.gambino@hsanmartino.it)

Laureato magistrale in Biostatistica presso l'Università degli Studi di Milano-Bicocca, dal 2023 è Biostatistico presso l'UOC Epidemiologia Clinica del San Martino dove si occupa della gestione delle qualità dei dati clinici al fine di preparare dataset per analisi di Intelligenza Artificiale e dell'utilizzo della piattaforma REDCap nell'ambito della creazione di e-CRF.





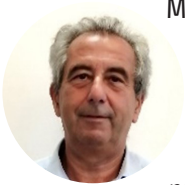
**Luisa Franco** [luisa.franco@hsanmartino.it](mailto:luisa.franco@hsanmartino.it)

Laureata in Biologia all'Università di Genova; riceve il titolo di PhD in Biotecnologie cellulari all'Università di Milano. Svolge attività di ricerca come postdoc presso University of Minnesota (USA), le Università Federico II di Napoli e di Genova e l'Istituto Gaslini; pubblica 41 articoli, H Index 29. Dal 2006 si occupa di gestione di ricerca, prima all'IIT come Neurofamily Coordinator e dal 2022 come dirigente tecnico/scientifico presso la Direzione Scientifica del San Martino.



**Cristina Grandi** [cristina.grandi@hsanmartino.it](mailto:cristina.grandi@hsanmartino.it)

Laureata in Giurisprudenza all'Università di Genova e abilitata all'esercizio della professione forense. Funzionario giuridico in Regione Liguria, Dipartimento Salute, con incarico di posizione organizzativa sulla Ricerca sanitaria. Dal 2020 dirigente amministrativo con alta professionalità presso la Direzione Scientifica del San Martino. I ruoli di responsabilità ricoperti sia in ente di programmazione che di gestione hanno determinato l'acquisizione di competenze specialistiche complete.



**Marco Damonte Prioli** [direzione.generale@hsanmartino.it](mailto:direzione.generale@hsanmartino.it)

Informatico ed economista, precedenti esperienze in qualità di Direttore Amministrativo e Generale, dal 2023 ricopre il ruolo di Direttore Generale dell'IRCCS Policlinico San Martino. Particolarmente sensibile alle tematiche della ricerca clinica, collabora assiduamente con la Direzione Scientifica del Policlinico, apportando le sue competenze giuridiche, economiche e di programmazione sanitaria, nonché dell'applicazione del codice di trattamento dei dati personali.



**Luca Boni** [luca.boni@hsanmartino.it](mailto:luca.boni@hsanmartino.it)

Laureato in Medicina e chirurgia e specializzato in Statistica sanitaria. Il suo percorso professionale si è sempre svolto nell'ambito della ricerca clinica, dove ha acquisito competenze in merito alle sue diverse fasi di realizzazione, dalla ideazione alla analisi e interpretazione dei risultati. Coautore di oltre 150 pubblicazioni scientifiche, fra articoli e libri, ha un H Index pari a 61 e attualmente è Direttore della U.O. Epidemiologia clinica dell'IRCCS Ospedale Policlinico San Martino.



**Antonio Uccelli** [direzione.scienfitica@hsanmartino.it](mailto:direzione.scienfitica@hsanmartino.it)

Direttore Scientifico dell'IRCCS Ospedale Policlinico San Martino, Professore ordinario di Neurologia. Coordinatore del Partenariato Esteso di Neuroscienze Mnesys e dal 2023 coordinatore scientifico del Progetto Bandiera della Regione Liguria. È co-autore di 258 articoli scientifici. Il suo H Index è 64. Il focus della sua ricerca sono le neuroscienze, le malattie autoimmuni e le cellule Staminali.

## **Il portale delle Raccolte Fotografiche: un progetto di condivisione dei patrimoni e delle conoscenze tra pubblico e privato**

Alessandro Coco

Ministero della Cultura ICCD

**Abstract.** Produzione di campagne fotografiche e descrizione e gestione del patrimonio culturale sono i due compiti che caratterizzano l'ICCD; il Portale delle Raccolte Fotografiche è il frutto del continuo affinamento dei sistemi di produzione delle immagini e dei dati descrittivi. Il Portale Fotografico Nazionale rappresenta un modello innovativo per la valorizzazione e la condivisione del patrimonio culturale italiano. Attraverso l'accesso aperto e la collaborazione, la piattaforma offre un importante strumento per la ricerca, l'educazione e la fruizione pubblica del patrimonio fotografico nazionale.

L'Istituto Centrale per il Catalogo e la Documentazione (ICCD) sin dalla sua nascita nel 1975 è impegnato nell'applicazione delle più avanzate tecnologie per la produzione e gestione della fotografia, fino alla fine del XX secolo nel formato analogico e da quasi trent'anni, nel formato digitale.

Forte anche della sua esperienza nella gestione degli archivi del Gabinetto Fotografico Nazionale, l'ICCD si adopera per la promozione della cultura fotografica e la valorizzazione di archivi e collezioni di fotografia.

In tale ambito, l'Istituto vanta la duplice competenza nella descrizione e gestione del patrimonio culturale (Catalogo) e nella produzione di campagne fotografiche (Documentazione).

- in merito alla descrizione e gestione del patrimonio culturale l'Ufficio del Catalogo cura il sistema del Catalogo Generale dei Beni Culturali, promuove e coordina iniziative nel settore della catalogazione, intesa come atto conoscitivo fondamentale per le pratiche di tutela, valorizzazione e fruizione del patrimonio culturale italiano.

- nell'ambito della produzione di campagne fotografiche: Il Gabinetto Fotografico Nazionale, operativo ininterrottamente dal 1895, rappresenta un punto di riferimento per le attività di documentazione fotografica del patrimonio culturale italiano. L'archivio delle fotografie realizzate e acquisite nel corso degli anni consta di più di cinquecento mila negativi e più di due milioni di positivi che assieme alle fotografie dell'Aerofototeca Nazionale ammonta a un patrimonio di circa dieci milioni di fotografie; il tema prevalente è la documentazione del patrimonio culturale italiano in tutti i suoi aspetti: storico artistico, architettonico, urbanistico e paesaggistico, immateriale.

Il Portale delle Raccolte Fotografiche nasce in tale contesto ed è il frutto di un continuo affinamento dei sistemi di produzione sia delle immagini che dei dati descrittivi. Ad oggi, sono rese disponibili al pubblico oltre quattrocentomila fotografie, in gran parte digita-

lizzate, degli archivi ICCD e di altri Istituti. Inoltre, il portale ha anche visto il coinvolgimento, in specifici progetti, delle più giovani generazioni di fotografi nativi digitali. Il loro contributo è stato fondamentale nella messa a punto dei sistemi di raccolta dei dati descrittivi, garantendo un approccio innovativo e al passo con le tecnologie più recenti.

La realizzazione del portale ha offerto all'ICCD l'opportunità di riorganizzare e strutturare non solo le collezioni e le attività, ma soprattutto le idee sulla fotografia che guidano il nostro lavoro nella promozione della cultura fotografica. La qualità dei dati, l'attenzione alla condivisione delle immagini, l'apertura e il riuso dei dati e delle immagini, il rigore descrittivo e la tutela dell'oggetto fotografico riflettono il nostro concetto di fotografia come bene culturale che guidato nel dare forma alle funzionalità del portale.

Il portale è alimentato da un sistema gestionale messo a punto da ICCD e basato sul software open source xDams, sviluppato dalla ditta Regesta. L'adozione di standard descrittivi internazionali, arricchiti per soddisfare specifiche esigenze descrittive ci ha permesso di mettere a punto un tracciato agile e di immediata usabilità. Il software è a disposizione delle istituzioni culturali che scelgono di aderire al portale, adottando il medesimo approccio promosso dall'ICCD. La pubblicazione sul portale consente agli istituti partecipanti di valorizzare il proprio patrimonio in una sezione dedicata.

Il portale è anche l'occasione per arricchire i vocabolari controllati e condivisi tra tutte le istituzioni partecipanti, per una più puntuale descrizione, in particolare, delle tecniche fotografiche e delle categorie descrittive; è stata anche avviata una rigorosa redazione dell'authority file dei fotografi, che grazie all'ampio arco cronologico coperto delle collezioni, permette di coprire l'intera storia della fotografia in particolare italiana.

Vi propongo una navigazione del sito web risponde alle idee che più di altre hanno modellato la forma e le funzionalità del portale e sono la declinazione del nostro concetto di fotografia come bene culturale: la fotografia elemento d'archivio, la fotografia strumento di promozione, la fotografia oggetto da tutelare, la fotografia bene da condividere.

## 1 - LA FOTOGRAFIA ELEMENTO D'ARCHIVIO

Ministero della Cultura

Raccolte Fotografiche

☰


Ministero della Cultura

Istituto centrale per il catalogo e la documentazione

☰


### Gli Istituti

4 elementi




Istituto Autonomo Villa Adriana e Villa d'Este

743 FOTOGRAFIE




Istituto centrale per il catalogo e la documentazione

404.339 FOTOGRAFIE



Soprintendenza Archeologia, Belle Arti e Paesaggio del Friuli Venezia Giulia

827 FOTOGRAFIE



Soprintendenza Archeologia, Belle Arti e Paesaggio per le province di Ravenna, Forlì-Cesena e Rimini

373 FOTOGRAFIE

### Guida al Patrimonio

RAGGRUPPAMENTO

- ▼ **Gabinetto Fotografico Nazionale**
  - FONDO
  - ▼ **Ministero della Pubblica Istruzione. Direzione generale Antichità e Belle Arti. Gabinetto Fotografico Nazionale**
    - 1895 - 1974
    - SERIE
    - Direzione Giovanni Gargioli**
    - 1895 - 1913
    - SERIE
    - Direzione Carlo Carboni**
    - 1913 - 1932
    - SERIE
    - Direzione Luigi Serra**
    - 1932 - 1938
    - SERIE
    - Direzione Bito Coppola**
    - 1938 - 1958
    - SERIE
    - Direzione Giorgio Castellfranco**
    - 1958 - 1964
    - SERIE
    - Direzione Carlo Bertelli**
    - 1964 - 1973
    - SUBSTANZO
    - **Fondi acquisiti dal Gabinetto Fotografico Nazionale**
    - 1895 - 1974
    - FONDO
    - **Ministero della Cultura. Istituto Centrale per il Catalogo e la Documentazione. Gabinetto Fotografico Nazionale**
    - 1975 -
    - FONDO
    - **Ministero della Pubblica Istruzione. Direzione generale Antichità e Belle Arti. Archivio fotografico**
    - 1863 - 1975
    - FONDO
    - **Ministero della Cultura. Istituto Centrale per il Catalogo e la Documentazione. Aerofototeca Nazionale**
    - 1969 -
    - FONDO
    - **Ministero della Cultura. Istituto Centrale per il Catalogo e la Documentazione. Progetti di fotografia contemporanea**
    - 2022 -

Il portale è progettato per ospitare non solo l'archivio dell'ICCD, ma anche archivi di fotografia di enti conservatori differenti.

Ciascun ente, sia esso del MiC o di altra natura giuridica, può pubblicare sul portale le proprie collezioni, avere accesso agli authority già presenti e arricchirli con nuovi record redatti secondo il tracciato condiviso.

Abbiamo deciso di mettere a disposizione dell'utente la Guida al patrimonio, quale strumento di navigazione utile ad avere un visione d'insieme dell'archivio e apprezzare i cambiamenti avvenuti all'interno di ciascun archivio, nel corso degli anni. I fondi sono disposti cronologicamente e rendono leggibili le varie fasi di acquisizioni e produzioni.

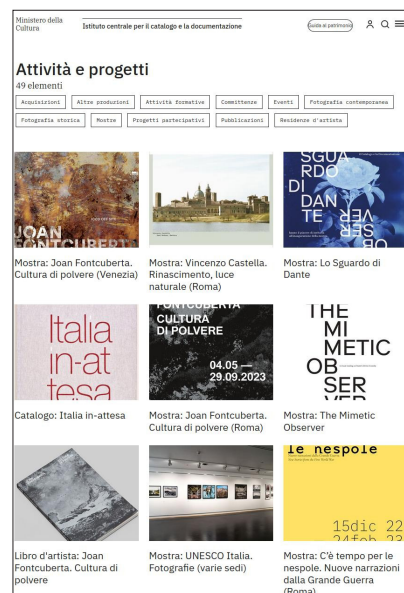
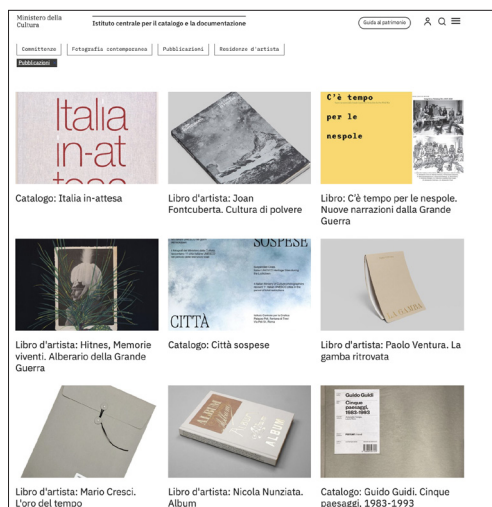
Dalla descrizione archivistica si può passare alla visualizzazione delle relative fotografie. L'attenzione alla descrizione archivistica è affiancata da un'accurata descrizione di ogni singola fotografia e supportata da authority file delle tecniche, degli autori dell'opera fotografata e non ultimo dei fotografi.

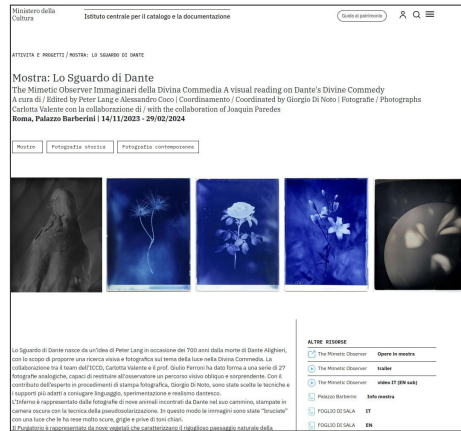
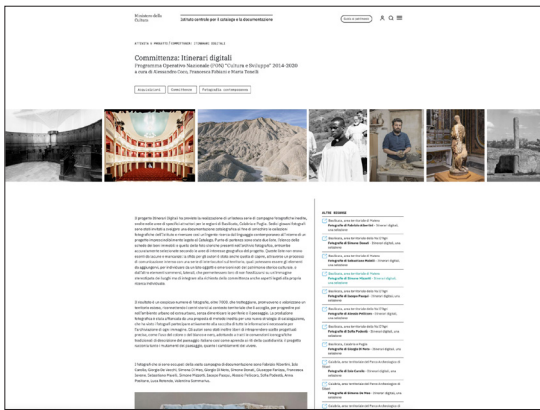
## 2 - LA FOTOGRAFIA STRUMENTO DI PROMOZIONE

La sezione Attività e progetti dà spazio alle strategie di promozione della fotografia che l'istituto persegue con l'organizzazione di mostre in Italia e all'estero delle proprie collezioni; con pubblicazioni di libri e cataloghi che raccolgono l'esito di committenze e residenze d'artista; con la promozione di eventi, workshop e convegni sui temi della fotografia storica e contemporanea e non ultimo con la collaborazione con scuole di fotografia per la produzione di campagne di documentazione. Una promozione esclusivamente on line è la pubblicazione di percorsi di lettura dell'archivio originali e in grado di creare connessione anche con gli altri sistemi informativi dell'ICCD.

## 3 - LA FOTOGRAFIA OGGETTO DA TUTELARE E CONSERVARE

Navigare le collezioni fotografiche dell'ICCD permette di attraversare anche la storia delle tecniche di stampa e di ripresa che hanno determinato i linguaggi fotografici. L'ICCD è quotidianamente impegnato nella salvaguardia della fotografia analogica nella sua dimensione materiale accompagnata dalla necessaria conoscenza delle tecniche fotografiche. Per





questo nel sito abbiamo previsto una sezione, curata dai nostri esperti, in cui si danno informazioni sulle caratteristiche salienti di ogni singola tecnica. Ciò ci ha permesso di pubblicare un rigoroso authority, collegato ai thesaurus internazionali, e reso disponibile come tutti i dati del sito in formato LOD.



Le collezioni fotografiche pubblicate sul portale sono riutilizzabile in differenti modalità. Ciascuna di esse è una sorta di rimodulazione del contesto della singola immagine. Il grado minimo di riuso possibile è una sorta di slittamento interno: l'utente può creare delle selezioni di immagini e pubblicare le raccolte così create, in una pagina dedicata sul portale, rendendo le fotografie, accessibili non più secondo il loro ordinamento archivistico originale, ma con criteri personali non meno interessanti che possono innescare letture inedite e stimolanti del patrimonio.



Un livello differente di riuso è garantito dall'uso del IIIF International Image Interoperability Framework: i dati descrittivi e le immagini a piena risoluzione possono essere usate fianco a fianco ad oggetti digitali provenienti da altri ecosistemi informativi che hanno adottato il IIIF. Le immagini non solo non sono più nella loro posizione archivistica ma possono essere

così accessibili anche su altre piattaforme.

Il modo più generico per mettere l'utente nelle condizioni di riusare le fotografie è il download libero che, seppur non è l'opzione migliore, perché non garantisce la circolazione dei dati assieme alla fotografia, è una scelta importante da parte dell'istituto in termini di accesso aperto alle risorse del Ministero della Cultura.

Il livello più avanzato di riuso dei dataset intesi come dati descrittivi e immagini digitali è l'uso dei Linked Open Data resi disponibili tramite l'endpoint Sparql Virtuoso sul sito [dati.cultura.gov.it/sparql](http://dati.cultura.gov.it/sparql)

## Conclusioni

L'Istituto Centrale per il Catalogo e la Documentazione (ICCD) ha consolidato il proprio ruolo di promotore e custode della fotografia come bene culturale. Attraverso il Portale delle Raccolte Fotografiche, l'ICCD ha reso accessibile un vasto patrimonio di immagini e dati, garantendo qualità e innovazione grazie a strumenti come il IIIF e i Linked Open Data. L'istituto continua a promuovere la fotografia, non solo come documento storico, ma come risorsa condivisibile e fruibile, confermando il suo impegno nella valorizzazione e tutela del patrimonio culturale italiano.

## Autore

Alessandro Coco [alessandro.coco@cultura.gov.it](mailto:alessandro.coco@cultura.gov.it)

Laureato in Scienze della Comunicazione a Roma e in Arti Visive allo IUAV, ha conseguito un dottorato a Udine sullo stile documentario. È record manager e coordinatore del portale delle raccolte fotografiche dell'ICCD. È tra gli ideatori del portale #scenedaunpatrimonio per la fotografia vernacolare; ha contribuito alle linee guida per la digitalizzazione dell'Istituto Centrale per la Digital Library. Cura mostre fotografiche per l'ICCD e altre istituzioni. Fa parte del gruppo di ricerca visiva Studio Bellosguardo in cui si occupa della progettazione e gestione di archivi digitali partecipati.

