



# LA MEMORIA DIGITALE

FORME DEL TESTO E ORGANIZZAZIONE DELLA CONOSCENZA  
ATTI DEL XII CONVEGNO ANNUALE AIUCD

UNIVERSITÀ DI SIENA, 5-7 GIUGNO 2023

ASSOCIAZIONE PER  
L'INFORMATICA UMANISTICA  
E LA CULTURA DIGITALE



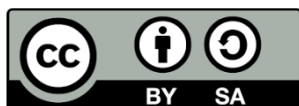
ISBN 978-88-942535-7-3

Copyright ©2023 AIUCD  
Associazione per l'Informatica Umanistica e la Cultura Digitale



Il presente volume e tutti i contributi sono rilasciati sotto licenza  
Creative Commons Attribution Share-Alike 4.0 International license ([CC-BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)).  
Ogni altro diritto rimane in capo ai singoli autori.

This volume and all contributions are released under the  
Creative Commons Attribution Share-Alike 4.0 International license ([CC-BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)).  
All other rights retained by the legal owners.



A cura di: Carbé, Emmanuela ; Lo Piccolo, Gabriele ; Valenti, Alessia; Stella, Francesco (2023). La memoria digitale: forme del testo e organizzazione della conoscenza. Atti del XII Convegno Annuale AIUCD, Siena: Università degli Studi di Siena

Ogni link citato era attivo al 22 maggio 2023, salvo ove diversamente indicato.  
All links have been visited on 22th May 2023, unless otherwise indicated

Si prega di notificare all'editore ogni omissione o errore si riscontri, al fine di provvedere alla rettifica.  
Please notify the publisher of any omissions or errors found, in order to rectify them.  
**aiucd.segreteria [at] aiucd.org**

I contributi pubblicati nel presente volume hanno ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione anonima mediante *double-blind peer review* sotto la responsabilità del Comitato Scientifico di AIUCD 2023.

All the papers published in this volume have received favourable reviews by experts in the field of DH, through an anonymous double-blind peer review process under the responsibility of the AIUCD 2023 Scientific Committee.

Il programma della conferenza AIUCD 2023 è disponibile online

The AIUCD 2023 Conference Program is available online

<http://www.aiucd2023.unisi.it>

### **Comitato Scientifico**

Nicola Barbuti  
Marina Buzzoni  
Emmanuela Carbé (co-chair)  
Fabio Ciraci  
Fabio Ciotti  
Angelo Mario Del Grosso  
Maurizio Lana  
Monica Monachini  
Paolo Monella  
Roberto Rosselli Del Turco  
Gino Roncaglia  
Francesco Stella (chair)  
Francesca Tomasi

### **Comitato organizzativo**

Elisabetta Bartoli  
Paola Bellomi  
Monica Bianchini  
Silvia Calamai  
Riccardo Castellana  
Rosalba Nodari  
Antonio Rizzo  
Enrico Zanini

**Segreteria del Convegno:** Gabriele Lo Piccolo, Francesca Pietrini;

Giulia Bassi, Silvia Cappa, Chiara Cauzzi, Martina Corti, Elena Crocicchia, Anna Guadagnoli, Giada Giannetti, Bogdan Groza, Alessia Luvisotto, Paola Mocella, Pietro Orlandi, Martina Paccara, Elisa Petri, Maria Grazia Schiaroli

Si ringrazia l'Ufficio stampa, comunicazione istituzionale e stampa digitale, l'Ufficio Ricerca, Biblioteche, Internazionalizzazione e Terza Missione, e il Supporto eventi culturali e convegnistici dell'Università di Siena

Supporto tecnico: MCM Service, Santa Chiara Lab, Presidio San Niccolò

### **Enti organizzatori**

AIUCD;

Università degli Studi di Siena: Dipartimento di Filologia e critica delle letterature antiche e moderne (DFCLAM), Centro interuniversitario di Studi Comparati I Deug-Su, Centro Interdipartimentale di Ricerca Franco Fortini in "Storia della tradizione culturale del Novecento", Santa Chiara Lab, in collaborazione con i Dipartimenti di Ingegneria dell'Informazione e Scienze Matematiche (DIISM), di Scienze sociali, politiche e cognitive (DISPOC) e di Scienze storiche e dei beni culturali (DSSBC);

CLARIN-IT.

**Con il patrocinio di:** Journal of the Text Encoding Initiative

## **Chair di track**

### **Archivi, edizioni digitali, organizzazione della conoscenza**

Marina Buzzoni, Paolo Monella, Roberto Rosselli Del Turco

### **Analisi computazionale dei testi**

Fabio Ciotti, Rachele Sprugnoli

### **Intelligenza Artificiale e modelli applicati ai beni culturali**

Monica Bianchini, Federico Boschetti

### **Preservazione della memoria e del patrimonio digitale**

Nicola Barbuti, Maurizio Lana

### **Workshop**

Francesco Stella, Emmanuela Carbé

## **Lista dei revisori**

Stefano Allegrezza, Cristiano Amendola, Paolo Andreini, Laura Antonietti, Luca Bandirali, Sofia Baroncini, Elisabetta Bartoli, Stefano Bazzaco, Andrea Bellandi, Paola Bellomi, Benedetta Bessi, Andrea Bolioli, Luca Bombardieri, Simone Bonechi, Alice Borgna, Flavia Bruni, Paolo Buono, Dino Buzzetti †, Silvia Calamai, Anna Cappellotto, Giuliana Capriolo, Vittore Casarosa, Riccardo Castellana, Paola Castellucci, Simona Chiodo, Fabio Ciraci, Elisa Corrà, Elisa Cugliana, Fabio Cusimano, Christian D'Agata, Elisa D'Argenio, Vincenza D'Urso, Stefano Dall'Aglia, Marilena Daquino, Angelo Mario Del Grosso, Antonio Di Silvestro, Diego Mantoan, Dominique Brunato, Dominique Longrée, Edmondo Grassi, Elena Spadini, Giulia Fabbris, Pierluigi Felicciati, Paolo Fioretti, Franz Fischer, Greta Franzini, Francesca Frontini, Daniele Fusi, Simone Giusti, Marco Grasso, Fabiana Guernaccini, Alessandro Iannella, Benedetta Iavarone, Alessandro Lenci, Eleonora Litta, Agnese Macchiarelli, Marco Maggini, Elisabetta Magnanti, Francesco Mambrini, Tiziana Mancinelli, Anna Maria Marras, Cristina Marras, Luca Martinelli, Stefano Melacci, Federico Meschini, Alessio Miaschi, Andrea Micheletti, Giovanni Morrone, Rosalba Nodari, Giuseppe Palazzolo, Niccolò Pancino, Fiammetta Papi, Enrico Pasini, Marco Passarotti, Giulia Pedonese, Igor Pizzirusso, Federico Ponchio, Francesca Pratesi, Alessia Lucia Prete, Simone Rebor, Giulia Renda, Gino Roncaglia, Irene Russo, Enrica Salvatori, Eva Sassolini, Daniele Silvi, Daria Spampinato, Linda Spinazze', Francesco Stella, Matteo Tiezzi, Francesca Tomasi, Sara Tonelli, Gennaro Vessio, Paul Gabriele Weston, Michelangelo Zaccarello, Patrizia Zambrano, Marco Zappatore, Andrea Zugarini

# LA MEMORIA DIGITALE

## XII CONVEGNO ANNUALE AIUCD

Siena, 5-7 giugno 2023

### Indice

|                                                                                                                                                             |          |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|
| <i>Prefazione</i>                                                                                                                                           | IV       |
| <i>In ricordo di Dino</i>                                                                                                                                   | VII      |
| <b>1. Archivi, edizioni digitali, organizzazione della conoscenza</b>                                                                                       | <b>1</b> |
| Creare e interrogare i modelli: un duplice valore conoscitivo. Il caso dei pareri di lettura einaudiani di Natalia Ginzburg                                 | 2        |
| Exploring iconographical and iconological content in semantic data through art historians' interpretations                                                  | 9        |
| L'applicazione del riconoscimento testi neurale per la realizzazione di ristampe digitali                                                                   | 15       |
| PATH, un thesaurus per le arti performative. Osservazioni empiriche e criticità d'uso                                                                       | 24       |
| I dati della ricerca: gestione, diffusione e preservazione nei Next generation repositories                                                                 | 28       |
| ODI and BACODI: a study on <i>Destini incrociati</i> by Italo Calvino with Semantic Web Technologies                                                        | 35       |
| L'edizione di un graphic novel medievale: il <i>Romanzo di Tristan</i> in München, BSB, Cgm 51                                                              | 41       |
| I <i>Capricci d'un Re</i> digitale                                                                                                                          | 48       |
| Harmonizing and modelling a bibliographic ontology of imperial diplomas within the OntoVE Knowledge Base                                                    | 51       |
| Mappare il museo in IIIF. Una combinazione di deep zoom e VR360 per la Galleria Borghese di Roma                                                            | 57       |
| H2IOSC: Humanities and Heritage Open Science Cloud                                                                                                          | 63       |
| Śivadharmā Database CMS. HTML and graph as a starting point for digital editions                                                                            | 65       |
| L'approccio human-centered nella progettazione di un ecosistema digitale che massimizzi la diffusione del materiale diaristico preservandone la singolarità | 71       |
| Un futuro digitale per gli archivi de 'Gli Anni della Cupola'                                                                                               | 78       |
| <i>Ianus 2.0</i> . Corrottele filologiche, congetture digitali                                                                                              | 85       |
| Verso una «volontà d'archivio» digitale                                                                                                                     | 92       |
| L'applicazione del digitale alla valorizzazione della lingua come patrimonio culturale immateriale                                                          | 99       |
| Digital <i>Accrocchio</i> : a computational image searching tool for social history                                                                         | 104      |
| Quali prospettive per ItaDraCor? Risorse e strumenti per la codifica di testi drammatici in lingua italiana                                                 | 108      |
| Metodologie di indagine per una nuova piattaforma: Digital Scholarly Editions Platform and Aligned Translation                                              | 111      |
| Making Korean Perseus Digital Library: The Case Study of 'Dokdo Digital Library'                                                                            | 114      |
| Constructing an Old English WordNet: The Case of <i>Guilt</i>                                                                                               | 122      |
| A digital diplomatic edition of the 10th-century charters of Lucca for Latin corpus linguistics                                                             | 124      |
| On why and how we should build a catalogue of software products for digital scholarly editing                                                               | 130      |
| Database Allegorico Dantesco                                                                                                                                | 134      |
| Tecnologie digitali per la realizzazione e adattamento di contenuti scolastici facilitati e semplificati                                                    | 139      |
| <i>Proteus</i> : marcatura di testi per una lettura antropologica del mito greco e romano                                                                   | 145      |
| Seeing the Unseen: an approach for visualising situated statements in the Cultural Heritage domain                                                          | 149      |
| <i>Soavissime ariette</i> : impiego di repertori digitali per lo studio della tradizione delle arie d'opera                                                 | 154      |
| <i>Musisque Deoque</i> : la nuova organizzazione dei testimoni manoscritti                                                                                  | 159      |
| Storytelling with Linked Open Data                                                                                                                          | 164      |
| From ontology design to user-centred interfaces for music heritage                                                                                          | 168      |
| Un modello open data per l'interazione utente: la digital library Open Memory Apulia                                                                        | 173      |
| Nel laboratorio di <i>Poeti d'Italia in lingua latina</i> : lavori in corso e prospettive future                                                            | 177      |

|                                                                                                                                             |     |
|---------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Metadata e concetti, una nuova intermediazione delle fonti. Quali prospettive per la ricerca e la diffusione della storia?                  | 181 |
| Metodi di armonizzazione per codifiche TEI cooperative: l'edizione Bellini Digital Correspondence                                           | 185 |
| LINEA: per un portale della letteratura italiana del Novecento                                                                              | 191 |
| ‘Voce dei Margini’: Le annotazioni di Giorgio Bassani                                                                                       | 197 |
| Scholarly edition of born-digital literary sources: a case study                                                                            | 204 |
| Per un atlante delle biblioteche medievali: il progetto Mirabile-Atlas e una nuova prospettiva sulla circolazione della cultura manoscritta | 208 |
| Dai metadati ai dati, dai contesti ai contenuti: aumentare la descrizione archivistica                                                      | 214 |
| ITINERA: progetto, database e prospettive                                                                                                   | 220 |
| Per l'edizione digitale delle lettere familiari di Verga: un esperimento di analisi semantica                                               | 224 |
| Ease the collaboration making Scholarly Editions: the GreekSchools case study                                                               | 230 |

## 2. Analisi computazionale dei testi 233

|                                                                                                                                                                                          |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Zoo-KG: A knowledge graph of zoological data                                                                                                                                             | 234 |
| <i>Do AIs dream of electric comics?</i> Generative AI models, digital memory, and creativity                                                                                             | 237 |
| ChatGPT: un Pappagallo Stocastico può essere di aiuto a un Vero Ricercatore (Umanistico)?                                                                                                | 245 |
| Un atlante digitale per i borghi delle aree interne                                                                                                                                      | 251 |
| To what extent are lemmatisation and annotation relevant for deep learning assignments and textual motifs detection? The case-study of Peter Damian's letters (11 <sup>th</sup> century) | 254 |

## 3. Intelligenza artificiale e modelli applicati ai beni culturali 260

|                                                                            |     |
|----------------------------------------------------------------------------|-----|
| <i>Falso movimento</i> (Moretti 2022): Memories and Translations           | 261 |
| Verso la creazione della versione italiana di SPARSAR                      | 265 |
| Unsupervised Information Extraction from Medieval Charters                 | 275 |
| The challenges of sense granularity in word sense disambiguation for Latin | 279 |
| Machine Learning Algorithm for the Scansion of Old Saxon Poetry            | 284 |
| Diachronic and diatopic word embeddings from British historical newspapers | 289 |
| GPT-3 vs. Delta. Applying stylometry to large language models              | 292 |
| EcoLet. Ecotemario della Letteratura italiana                              | 298 |
| Risorse linguistiche per lo studio dei <i>Promessi Sposi</i>               | 301 |

## 4. Preservazione della memoria e del patrimonio digitale 304

|                                                                                                                           |     |
|---------------------------------------------------------------------------------------------------------------------------|-----|
| Uso e riuso di dati per la valorizzazione del patrimonio culturale italiano online: buone pratiche per l'interoperabilità | 305 |
| “Dalla civiltà dell'archivio alla civiltà dell'algoritmo?”                                                                | 309 |
| <i>Breaking the Canvas</i> A Serious Game to challenge stereotypes in Visual Culture through Fine Arts                    | 315 |
| CoMOnTology. Towards An Ontology for the Chronology of Mediterranean archaeologies: a model for the digital memory        | 322 |
| Dematerializzazione e sostenibilità: prospettive ideali e problemi concreti                                               | 327 |
| New Media Art: indagini conservative sulle eredità immateriali dell'arte di fine millennio                                | 332 |
| The challenges of preservation: digitizing graffiti in the urban landscape                                                | 338 |
| The Lions of Venice Arsenal: a neuroscientific approach to Cultural Heritage                                              | 343 |
| L'immaginazione tra 0 e 1. Alcune considerazioni su memoria e oblio nell'era digitale                                     | 347 |
| A methodological study on the projects developed by the Cologne Center for Humanities (CCeH)                              | 351 |
| Preservare il <i>Tesoro</i> . Per un nuovo sistema di interrogazione dei corpora dell'Opera del Vocabolario Italiano      | 358 |
| La galassia MQDQ tra innovazione e tradizione                                                                             | 365 |
| Memorizzare o ricordare? La memoria come fatto emozionale                                                                 | 371 |

|                                                                                                                               |         |
|-------------------------------------------------------------------------------------------------------------------------------|---------|
| GECA 3.0 – A new tool for cataloguing and enjoying cultural heritage                                                          | 373     |
| Libri, memoria e territorio: la digital library di Carlo Alberto                                                              | 380     |
| Un modello di catalogazione digitale delle grammatiche bilingue di spagnolo per italofoni                                     | 382     |
| Un'ontologia per la classificazione semantica dei dizionari storici: l'esempio del LEI Digitale                               | 385     |
| How many oral archives are in your home? Piloting a new Tuscan census in the Gra.fo Reloaded project                          | 389     |
| Le mafie e le Digital Humanities                                                                                              | 395     |
| I Confini della Lunigiana: un progetto di Digital Public History                                                              | 401     |
| L'immaginario francese della Puglia: una collezione digitale del 'Petit Tour' dell'altra Italia                               | 404     |
| <br><b>5. Workshop</b>                                                                                                        | <br>409 |
| Workshop sulle edizioni digitali: preparazione con codifica XML TEI e visualizzazione con il software EVT                     | 410     |
| Trascrivere, analizzare e pubblicare documenti d'archivio: Transkribus e l'intelligenza artificiale al servizio dello storico | 416     |
| Codificare (al)l'Infinito                                                                                                     | 421     |

## Prefazione

La dodicesima edizione del Convegno annuale dell'Associazione per l'Informatica umanistica e la cultura digitale (AIUCD) è intitolata *La memoria digitale: forme del testo e organizzazione della conoscenza*, con un programma che ha l'intento di interrogarsi sulle forme della testualità, intese in senso ampio e trasversale, dando spazio a riflessioni teoriche e metodologiche negli ambiti della filologia digitale, dell'analisi computazionale, dell'organizzazione e preservazione di collezioni digitali e nate digitalmente, dell'applicazione di tecniche di Intelligenza Artificiale e dei problemi della conservazione del patrimonio digitale.

La sede di Siena ha ricevuto il testimone a Lecce, durante l'undicesimo Convegno annuale che si è tenuto presso l'Università del Salento dall'1 al 3 giugno 2022: un passaggio di consegne che ci onora e che cade nel quindicesimo anno dalla fondazione del Master in Informatica del testo ed edizione elettronica (Infotext), oggi diretto da Elisabetta Bartoli, e a venticinque anni dal pionieristico avvio delle attività digitali del Centro di Studi Comparati I Deug-Su. Quest'anno il Dipartimento di Filologia e critica delle letterature antiche e moderne dell'Università di Siena è stato inoltre riconfermato Dipartimento di Eccellenza: il progetto del quinquennio 2023-2027 include un rafforzamento delle attività di ricerca legate all'Informatica umanistica, con la prosecuzione delle numerose collaborazioni con altri dipartimenti dell'Ateneo e con enti nazionali e internazionali.

Nel 2022 Fabio Ciracì e Patrizia Miggiano avevano proposto un ricchissimo programma di incontri online, "Loading AIUCD2022", per prepararci insieme al convegno di Lecce. Abbiamo idealmente proseguito l'iniziativa con "Aspettando AIUCD2023", un ciclo di workshop online organizzato in collaborazione con CLARIN-IT e con il suo Knowledge Centre for Digital and Public Textual Scholarship, con il sostegno del Centro di Ricerca Interdipartimentale in Digital Humanities dell'Università del Salento, del *Journal of the Text Encoding Initiative* e di IIF. "Aspettando AIUCD2023" ha avuto come docenti dei workshop Francesco Mambrini (*Introduzione a Python per l'analisi dei dati testuali*, 4 maggio 2023), Linda Spinazzè (*Zotero per la gestione dei dati bibliografici e la ricerca collaborativa*, 11 maggio 2023), Angelo Mario Del Grosso (*Selezione, elaborazione e presentazione di documenti XML-TEI mediante i linguaggi XPath e XSL*, 18 maggio 2023), Tiziana Mancinelli (*Introduzione a IIF – International Image Interoperability Framework*, 25 maggio 2023) e Rachele Sprugnoli (*Introduzione al Natural Language Processing*, 1 giugno 2023). Sono state inoltre proposte due lezioni online aperte al pubblico, in collaborazione con il Master Infotext: Klaus Kempf,

*Ritorno al passato. Mostre virtuali e Iconic Turn* (13 marzo 2023), e Franco Moretti, *Semantica delle reti drammatiche* (2 maggio 2023).

La Call for Papers ha ricevuto 96 proposte da oltre duecento autrici e autori, provenienti dall'Italia e da Austria, Belgio, Brasile, Repubblica Ceca, Finlandia, Francia, Germania, Norvegia, Corea del Sud, Spagna, Svizzera, Regno Unito e Stati Uniti. La sezione "Archivi, edizioni digitali, organizzazione della conoscenza", presieduta da Marina Buzzoni, Paolo Monella e Roberto Rosselli Del Turco, ha ricevuto 47 proposte; la seconda, "Analisi computazionale dei testi", presieduta da Fabio Ciotti e Rachele Sprugnoli, 6 proposte; la terza, "Intelligenza Artificiale e modelli applicati ai beni culturali", presieduta da Monica Bianchini e Federico Boschetti, 14 proposte; la quarta, "Preservazione della memoria e del patrimonio digitale", presieduta da Nicola Barbuti e Maurizio Lana, 26 proposte; i Workshop, presieduti da chi scrive, hanno ricevuto 3 proposte, sottoposte a revisione come le quattro track della call. Sono stati accettati complessivamente 82 interventi, con una media di accettazione dell'85,4%: 47 comunicazioni orali, 32 poster, molti dei quali già disponibili su Zenodo e Figshare, e 3 workshop.

In linea con la Call for Papers e con i temi di questa edizione del convegno, il programma prevede relazioni di Keynote speaker: il 5 giugno sarà interamente dedicato alle infrastrutture di ricerca, con una sessione presieduta da Marina Buzzoni e Cristina Marras, e interventi di Lino Leonardi e Emiliano Degl'Innocenti (*Filologia digitale e infrastrutture di ricerca, nel PNRR e oltre*), Arianna Ciula (*Infrastrutture di ricerca: dal ruolo della modellizzazione nell'ecosistema di laboratorio*), Monica Monachini e Francesca Frontini (*Scienza aperta e formazione nelle infrastrutture di ricerca europee e nazionali per le risorse linguistiche*). Il secondo giorno sarà dedicato alla filologia digitale, con l'intervento di Paola Italia, introdotta da Roberto Rosselli Del Turco, *Edizione, Annotazione, Visualizzazione. Problemi e prospettive della/per la filologia digitale*. Il 7 giugno avremo infine il contributo di Thorsten Ries, introdotto da Francesca Tomasi, con una riflessione sul trattamento degli archivi nati digitalmente: *Born-digital Literary Archives – the Present and the Future of an Emerging Field*.

Desideriamo ringraziare i Chair di track, il Comitato scientifico e i revisori che hanno generosamente contribuito, tutti coloro che hanno partecipato alla Call for Papers e i numerosi iscritti al convegno. Un ringraziamento va ai chair di sessione che riuniranno in percorsi trasversali gli interventi: Nicola Barbuti, Paola Bellomi, Andrea Bolioli, Federico Boschetti, Flavia Bruni, Paola Castellucci, Fabio Ciraci, Angelo Mario Del Grosso, Giulia Giovani, Maurizio Lana, Paolo Monella, Daria Spampinato, Rachele Sprugnoli. Ringraziamo il Comitato Scientifico e il Comitato Organizzatore per il loro essenziale contributo, il Direttivo di AIUCD per la costante collaborazione, CLARIN-IT per il generoso supporto. Un ringraziamento speciale va alla Segreteria Amministrativa del DFCLAM, e in particolare a Francesca Pietrini; a Gabriele Lo Piccolo, Alessia Valenti, e a tutte e tutti i giovani che hanno collaborato e collaboreranno nei giorni della conferenza: Giulia Bassi, Silvia Cappa, Chiara Cauzzi, Martina Corti, Elena Crocicchia, Anna Guadagnoli, Giada Giannetti, Bogdan Groza, Alessia Luvisotto, Paola Mocella, Pietro Orlandi, Martina Paccara, Elisa Petri, Maria Grazia Schiaroli.

Questo convegno sarà particolarmente difficile, il nostro pensiero è infatti rivolto a Dino Buzzetti: la sua improvvisa scomparsa, il 23 aprile 2023, ha profondamente toccato la nostra Associazione. Ma sarà anche un momento che riunirà le socie e i soci per confermare l'impegno di tutti noi nell'operare ricordandoci il suo insegnamento, il suo rigore scientifico e il suo spirito di servizio uniti all'estrema generosità e gentilezza. Con il Direttivo di AIUCD desideriamo che questi giorni siano prima di tutto un'occasione per ricordare con profonda riconoscenza il nostro Presidente onorario, tra i fondatori dell'Associazione nel 2011: il 5 giugno, durante la cerimonia di inaugurazione, Francesca Tomasi ci

consegnerà un ricordo che abbiamo anticipato in questo volume; il 6 giugno interverranno Daniela Calanca e Claudia Buzzetti, insieme ad Anna Maria Tammaro; infine, in chiusura di convegno, avremo un contributo di Fabio Ciotti, a cui idealmente uniamo i numerosi messaggi di gratitudine pubblicati in queste settimane sul blog dell'Associazione.

Emmanuela Carbé

Francesco Stella

# I dati della ricerca: gestione, diffusione e preservazione nei Next generation repositories

Ilaria Belvedere<sup>1</sup>, Simona Turbanti<sup>2</sup>

<sup>1</sup> Biblioteca Tiziano Terzani, Campi Bisenzio (Firenze), Italia – [ilaria.belvedere9@gmail.com](mailto:ilaria.belvedere9@gmail.com)

<sup>2</sup> Università di Pisa, Italia – [sturbanti@gmail.com](mailto:sturbanti@gmail.com)

## ABSTRACT

Data la crescente varietà dei dati prodotti dalle comunità di ricerca, la gestione di queste informazioni “granulari” non può prescindere dall’implementazione di piattaforme in grado di garantire la loro conservazione, visibilità e riuso, orientate verso il modello dei Next generation repositories. Tra le soluzioni disponibili, il software open source Dataverse, pensato specificatamente per i dati della ricerca, permette di strutturare i contenuti in archivi virtuali all’interno dei quali i dati possono essere gestiti a tutto tondo, grazie alla disponibilità di metadati descrittivi generali e specifici, di un sistema codificato di citazione e di formati diversi di visualizzazione ed esportazione. Un’analisi condotta presso l’Università di Pisa, attraverso una versione demo del software in locale denominata DataversePI, ha mostrato le potenzialità dello strumento e i vantaggi che porterebbe all’istituzione in termini di conservazione, preservazione, accesso e valorizzazione delle informazioni create durante tutte le fasi della ricerca. I Next generation repositories delle diverse istituzioni possono formare, inoltre, una rete distribuita in cui i diversi archivi interagiscono tra loro, pur mantenendo ognuno le proprie specificità.

## PAROLE CHIAVE

Dati della ricerca; data management plans; accesso aperto; archivi istituzionali; Dataverse

## 1. INTRODUZIONE<sup>1</sup>

Gli enti di ricerca si trovano, ormai da decenni, a dover mettere in atto strategie per la gestione dei dati della ricerca, ovvero di quei prodotti informativi generati «durante tutto il ciclo di ricerca [che] includono qualsiasi informazione in formato binario digitale [...] creata, immagazzinata, acceduta e restituita con l’uso della tecnologia dei computer» ([9]), comprendenti pertanto sia le pubblicazioni che i materiali non pubblicati.

Come sottolineato nelle *Linee guida sull’interoperabilità tecnica delle pubbliche amministrazioni*<sup>2</sup>, all’interno del Piano triennale per l’informatica nella Pubblica amministrazione, «i dati pubblici sono un bene comune e una risorsa del Paese in grado di produrre valore migliorando i servizi, creandone di innovativi e contribuendo a creare nuovi modelli di business, competenze e posti di lavoro»<sup>3</sup>. Una gestione più efficace dei dati pubblici, la creazione di servizi focalizzati sui cittadini agevolati dall’interoperabilità tra enti, il supporto all’attività imprenditoriale e alla ricerca scientifica rappresentano i principali obiettivi del Dipartimento per la trasformazione digitale del Governo italiano.

La pianificazione delle attività di *Research data management* (RDM) è necessaria anche in vista delle direttive internazionali in materia di gestione dei dati. Il Programma quadro dell’Unione europea per la ricerca e l’innovazione Horizon 2020<sup>4</sup> – in vigore per gli anni 2014-2020 – attraverso l’Open research data pilot<sup>5</sup> ha previsto l’obbligo dell’accesso aperto dove possibile, secondo il principio «as open as possible, as close as necessary» per i risultati della ricerca finanziati nell’ambito del programma pensati per la pubblicazione in riviste *peer-reviewed*: si prescrive, quindi, di depositare le pubblicazioni in archivi di dati della ricerca e di redigere i relativi *Data management plans* (DMPs).

Il documento FAIR *guiding principles* ([15]), elaborato nel 2016, presenta quattro importanti principi, che non rappresentano standard, ma raccomandazioni da seguire per una corretta gestione dei dati della ricerca: quest’ultimi dovrebbero essere resi rintracciabili, accessibili, interoperabili e riutilizzabili (*findable, accessible, interoperable*,

<sup>1</sup> Questo contributo prende spunto dalla tesi magistrale in Informatica umanistica di Ilaria Belvedere dal titolo *I Research Data Repositories nella gestione dei dati della ricerca: ipotesi di implementazione di un repository Dataverse per l’Università di Pisa* (relatore prof.ssa Simona Turbanti, correlatore prof. Paolo Milazzo), discussa presso l’Università di Pisa il 29 settembre 2022.

<sup>2</sup> Cfr. Italia, Presidenza del Consiglio dei ministri, Dipartimento per la trasformazione digitale, *Linee guida sull’interoperabilità tecnica delle pubbliche amministrazioni*. <https://docs.italia.it/italia/piano-triennale-ict/lg-modellointeroperabilita-docs/it/bozza/index.html>

<sup>3</sup> Cfr. <https://innovazione.gov.it/progetti/dati-e-interoperabilita/>

<sup>4</sup> Cfr. [https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-2020\\_en](https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-2020_en)

<sup>5</sup> Cfr. <https://www.openaire.eu/what-is-the-open-research-data-pilot>

reusable). I principi FAIR non potrebbero articolarsi senza la presenza di una policy sui dati della ricerca stilata dall'istituzione e senza il *Data management plan*, documento necessario al momento della presentazione di un progetto di ricerca.

Secondo quanto definito proprio all'interno del documento *Guidelines on FAIR data management in Horizon 2020* ([7]), i DMPs sono le descrizioni del ciclo di gestione dei dati della ricerca in uno specifico progetto di ricerca e costituiscono un elemento cardine per una buona gestione dei dati. Horizon Europe<sup>6</sup>, che fa seguito a Horizon 2020 per gli anni 2021-2027, ribadisce quanto previsto in precedenza puntando all'apertura di tutti i dati della ricerca.

## 2. GLI ARCHIVI ISTITUZIONALI DELLA RICERCA

Nonostante la letteratura presenti da oltre un ventennio gli innumerevoli apporti positivi degli archivi istituzionali della ricerca (Institutional repositories, IR) ([10]), spesso gli studiosi percepiscono il deposito dei propri “prodotti” – metadati e full text – all'interno di questi strumenti solo come un ulteriore obbligo cui assolvere ai fini degli esercizi valutativi locali e nazionali delle università/enti di ricerca senza cogliere le opportunità, anche a livello personale, che ne derivano, in termini di visibilità del proprio lavoro. L'agevole reperibilità in rete dei dati descrittivi e l'accesso al testo pieno del contributo, se possibile e nella forma eventualmente consentita dalle politiche degli editori (versione editoriale, preprint, postprint), consente, infatti, una maggiore circolazione dei risultati e, per giovani studiosi, la possibilità di attivare legami con ricercatori di altre realtà geografiche e di “fare rete”.

La crescente produzione di dati di diverse tipologie ha dettato, inoltre, la necessità di pensare a modelli nuovi di archivi che tengano conto della complessità di queste risorse. Nel 2017 è stato pubblicato il report ([11]) dei risultati ottenuti dal gruppo di lavoro Next Generation Repository Working Group della Confederation of Open Access Repositories (COAR)<sup>7</sup>, fondamento di un'infrastruttura distribuita e globalmente connessa a beneficio di ricerca, innovazione e collaborazione della comunità accademica. Uno degli elementi più importanti di questo programma consiste nel prevedere che gli IR conservino materiali variegati, così da permettere una condivisione di tipologie diverse di prodotti della ricerca e un loro riconoscimento.

I Next generation repository considerano sia i metadati sia i prodotti della ricerca come risorse web da identificare tramite URI, avvalendosi quindi dei vocabolari controllati propri del web semantico con un approccio pragmatico. Il report presenta sei principi guida e il modello concettuale alla base della costituzione dei NGR; i principi definiscono come essenziale la presenza di un sistema distribuito, cioè di una rete di repositories che eviti il controllo monopolistico della singola istituzione e, al contempo, garantisca la tutela delle specificità di ogni contesto, puntando all'accessibilità delle risorse ([10, p. 8]).

Anche a livello europeo si punta l'accento sull'interazione tra istituzioni; in particolare, per quanto riguarda il settore della ricerca, l'interoperabilità tra le infrastrutture è il focus del documento pubblicato dalla European Open Science Cloud (EOSC)<sup>8</sup>, EOSC interoperability framework report ([6]), frutto del lavoro di EOSC FAIR Working Group<sup>9</sup> e EOSC Architecture Working Group<sup>10</sup>: si sottolinea l'importanza dell'interoperabilità per la gestione dei dati della ricerca, distinguendo tra interoperabilità tecnica, semantica, organizzativa e legale. Attraverso uno studio teorico e pratico, condotto anche tramite interviste ad alcune comunità di ricerca appartenenti a EOSC, sono state individuate problematiche, necessità e raccomandazioni da seguire per ognuno dei quattro tipi di interoperabilità. Nel programma di lavoro di Horizon Europe per l'anno 2023-2024 vengono ribadite queste considerazioni.

## 3. I NEXT GENERATION REPOSITORIES: IL SOFTWARE OPEN SOURCE DATAVERSE

Tenendo presente il modello dei Next generation repositories, un ente di ricerca che voglia implementare un archivio pensato specificatamente per la gestione dei dati della ricerca deve considerare diversi fattori. Durante la fase di pianificazione, oltre all'architettura della piattaforma e le sue caratteristiche ([11]), è importante vagliare anche le funzionalità di gestione, preservazione e accessibilità dei contenuti e l'integrazione con servizi esterni. A questo proposito, il recente documento *The TRUST principles for digital repositories* ([3]) ha presentato alcuni principi guida per la configurazione dei repository per una gestione dei dati FAIR: i primi due riguardano la trasparenza in merito ai servizi offerti dal repository e al deposito dei dati, e la responsabilità collegate, il terzo è relativo ai bisogni dell'utenza, mentre il quarto e il quinto sono incentrati sulla preservazione dei dati nel lungo periodo e sugli aspetti tecnologici.

<sup>6</sup> Cfr. <https://horizoneurope.apre.it/>

<sup>7</sup> Cfr. <https://www.coar-repositories.org/>

<sup>8</sup> Cfr. <https://eosc-portal.eu/>

<sup>9</sup> Cfr. <https://eoscsecretariat.eu/working-groups/fair-working-group>

<sup>10</sup> Cfr. <https://eoscsecretariat.eu/working-groups/architecture-working-group>

In riferimento alla scelta del software, tra le opzioni open source disponibili, si annoverano Eprints<sup>11</sup> e Dspace<sup>12</sup> da una parte, validi anche se pensati prevalentemente per la gestione solo delle pubblicazioni, e soluzioni come InvenioRDM<sup>13</sup>, Haplo<sup>14</sup> e Dataverse<sup>15</sup>. Dataverse, in particolare, è stato scelto da diverse istituzioni internazionali e da alcune realtà italiane come l'Università degli studi di Milano e l'Istituto italiano di tecnologia (IIT).

Dataverse, nato nel 2006 dalla collaborazione tra l'Institute for Quantitative Social Science (IQSS) dell'Università di Harvard e altre istituzioni e giunto attualmente alla versione 5.13, si presenta come uno strumento per la gestione dei dati della ricerca che consente un uso in autonomia ai fruitori e senza una specifica connotazione disciplinare; la sua architettura si basa su Java Enterprise Edition (JEE)<sup>16</sup>. Grazie al suo design modulare il software consente di contribuire al codice, di personalizzare le installazioni secondo le necessità delle istituzioni e di avere un nucleo funzionale ([4]): dalla versione 4.8.6 è stato introdotto un *external tool framework* per permettere l'integrazione con programmi esterni senza intervenire nel *core code*, come in passato ([5]).

L'intento del progetto Dataverse consiste nell'offrire ai ricercatori uno spazio dove poter caricare i contenuti della ricerca che consenta di assicurare, da una parte, visibilità, sicurezza e garanzie di preservazione ai dati, dall'altra personalizzazione e autonomia di gestione per i professionisti della ricerca.

L'organizzazione dei contenuti, suddivisi in archivi virtuali chiamati *Dataverse collections* che, a loro volta, possono contenere altre collezioni o datasets, è intuitiva e facilmente gestibile sia tramite interfaccia che utilizzando le API. In Dataverse i dati della ricerca, conservati sul server istituzionale, sono visibili grazie a un sistema formale di citazione – a livello di file e dataset, in fase di pianificazione per le collezioni –, interoperabili (protocollo OAI-PMH<sup>17</sup>), ricercabili e accessibili grazie alle *features* relative ai metadati ([2]) ([14]), analizzabili attraverso strumenti specifici e facilmente integrabili all'interno di pagine web esterne. Per quel che concerne in particolare i metadati, Dataverse ne supporta l'inserimento sia a livello di collezioni, sia a livello dataset (fig. 1) e file.



Figura 7. Le funzionalità relative ai metadati livello dataset.

Il software prevede l'utilizzo di standard internazionali, come Dublin core (DC)<sup>18</sup> e Documentation data initiative (DDI)<sup>19</sup>, e si avvale della collaborazione con DataCite<sup>20</sup> e Google dataset search<sup>21</sup>. Inoltre, come osservato da Bantin ([1]), la pubblicazione dei dati e dei metadati avviene separatamente, garantendo la sicurezza dei dati. Altre importanti strumenti offerte da Dataverse riguardano:

- la gestione dei ruoli degli utenti in relazione alle collezioni e ai datasets;

<sup>11</sup> Cfr. [https://wiki.eprints.org/w/Main\\_Page](https://wiki.eprints.org/w/Main_Page)

<sup>12</sup> Cfr. <https://dspace.lyrasis.org/>

<sup>13</sup> Cfr. <https://inveniordm.web.cern.ch/>

<sup>14</sup> Cfr. <https://support.haplo.com/repository>

<sup>15</sup> Cfr. <https://dataverse.org/about>

<sup>16</sup> Cfr. <https://www.oracle.com/java/technologies/java-ee-glance.html>

<sup>17</sup> Cfr. [https://it.wikipedia.org/wiki/Open\\_Archives\\_Initiative\\_Protocol\\_for\\_Metadata\\_Harvesting](https://it.wikipedia.org/wiki/Open_Archives_Initiative_Protocol_for_Metadata_Harvesting)

<sup>18</sup> Cfr. <https://www.dublincore.org/>

<sup>19</sup> Cfr. <https://ddialliance.org/>

<sup>20</sup> Cfr. <https://datacite.org/>

<sup>21</sup> Cfr. <https://datasetsearch.research.google.com/>

- la scelta di licenze d'uso o la creazione di propri termini d'uso personalizzati,
- il supporto per i Big data;
- la preservazione dei dati;
- l'integrazione con servizi esterni, per esempio con piattaforme per l'analisi dei dati.

#### 4. I VANTAGGI DI DATAVERSE

Per avere un riscontro concreto sull'implementazione di un repository Dataverse, sono stati contattati 17 rappresentanti di istituzioni che hanno adottato il software e analizzato le fonti dirette di una istituzione, per un totale di 18 casi studio. L'obiettivo è stato individuare le motivazioni dietro alla loro scelta, riassunte in un istogramma e relativa tabella (fig. 2).

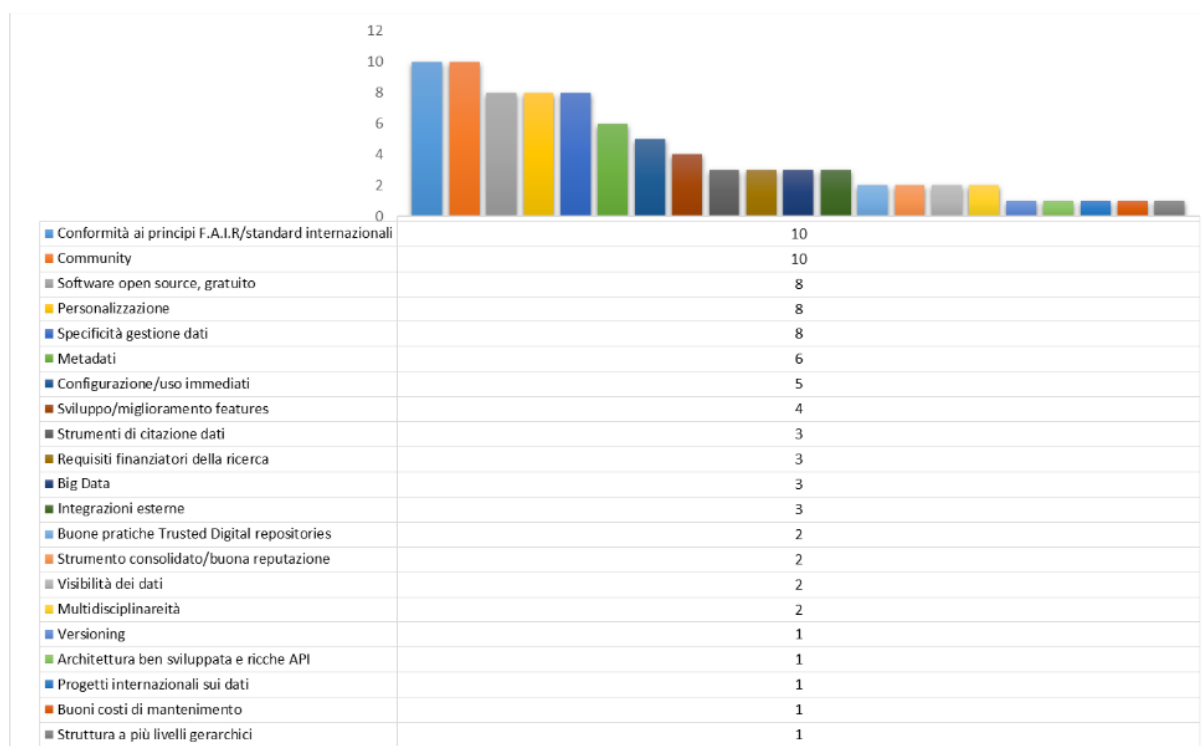


Figura 8. Feedback istituzioni.

Dalle interviste e dalle fonti è emerso che Dataverse potrebbe essere una buona soluzione per diverse ragioni. Le motivazioni più citate dalle istituzioni contattate sono, a pari merito, il fatto che Dataverse permetta una gestione dei dati della ricerca conforme ai principi FAIR e agli standard internazionali, il costante impegno della comunità di sviluppatori e contributori, oltre al supporto e disponibilità costanti verso gli utilizzatori del software. Gli altri tre vantaggi più menzionati riguardano la gratuità e la piena disposizione del codice, l'alto grado di personalizzazione in riferimento sia all'interfaccia, sia alla parte contenutistica e alla gestione dei ruoli e dei permessi, e la specificità, ovvero l'essere stato progettato in funzione dei *research data*. Un'altra importante ragione riportata da 6 istituzioni è l'attenzione ai metadati, la possibilità di creare blocchi di metadati personalizzabili, di utilizzare standard specifici e di effettuare l'*harvesting*. Nonostante la sua installazione e personalizzazione comportino certamente un costo di gestione, soprattutto in termini di risorse umane, le istituzioni coinvolte affermano che la configurazione e l'utilizzo del software sono immediati. Tra gli altri vantaggi segnalati, l'impegno della community nel migliorare le funzionalità già esistenti e nell'implementarne *ex-novo* e la presenza di strumenti di citazione specifici per i dati; peraltro, quest'ultimo elemento si inserisce nell'ambito della gestione dei dati e si collega alla già citata "visibilità" dei dati della ricerca. La comunità di Dataverse realizza collaborazioni in merito alla gestione dei dati della ricerca con fornitori di servizi analoghi, perseguendo l'interoperabilità, come chiaramente esplicitato nella Roadmap. In considerazione di ciò, il software può risultare un valido strumento per la realizzazione non soltanto di singoli repository di dati, ma anche di articolate infrastrutture di ricerca.

## 5. UN PROTOTIPO PER L'UNIVERSITÀ DI PISA: DATAVERSEPI

Al fine di testare direttamente le funzionalità di Dataverse è stato condotto uno studio, focalizzato sull'Università di Pisa, implementando una versione *demo* del software in locale, denominata DataversePI. Dopo l'installazione su una macchina virtuale di Ubuntu 18.04.6, è stato scaricato il pacchetto Archive in a box<sup>22</sup>, che consente una modalità di installazione immediata di Dataverse nella sua versione 5.11.1 avvalendosi del modulo Dataverse Docker, ideato da Data Archiving and Networked Services (DANS)<sup>23</sup>, Royal Netherlands Academy of Arts and Sciences (KNAW)<sup>24</sup> e Dutch Research Council (NWO)<sup>25</sup>. Archive in a box consente di servirsi delle funzioni più importanti di Dataverse e di avvalersi del supporto di vocabolari controllati esterni, di identificatori e DOI<sup>26</sup>, della personalizzazione di metadati e di altri strumenti offerti dal software nella sua versione classica.

La strutturazione dei contenuti ha previsto, a un primo livello, la *root collection*, pensata per contenere tutte le altre collezioni, a un secondo livello, le collezioni dei dipartimenti dell'Università di Pisa, le quali, a loro volta, ospitano le *collections* del livello successivo, corrispondenti agli archivi dei progetti di ricerca, dei gruppi di ricerca, dei laboratori di ricerca e dei ricercatori affiliati ai dipartimenti. Infine, a un quarto livello, sono presenti i dataset delle collezioni. All'interno del caso studio è stata creata la collezione del Dipartimento di filologia, letteratura e linguistica.

Per la gestione delle collezioni e dei loro contenuti sono state testate sia l'interfaccia sia la NATIVE API, un tipo di API di facile utilizzo anche per i non sviluppatori. La pagina della *root collection*, corrispondente alla homepage di DataversePI, presenta una maschera di ricerca *Google-like* e una funzione di ricerca avanzata. Sono disponibili, inoltre, filtri di visualizzazione e faccette di navigazione impostati sulla base dei metadati di citazione; oltre ai tre campi assegnati di default – Dataverse category (che indica la tipologia di Dataverse, per esempio Department), Publication year e Publication status – sono stati selezionati quelli ritenuti più significativi, ossia Subject, Keyword term, Type of data, Language, Grant information Grant agency, Grant information, Grant number e Deposit date.

Oltre ai metadati citazionali, le informazioni scelte per il Dipartimento di filologia, letteratura e linguistica sono Geospatial metadata, Social science and humanities Metadata e Journal metadata, in linea con le discipline studiate dai membri del Dipartimento; sono state, inoltre, mantenute le faccette di navigazione della *root collection*.

Il Dataverse FILELI ospita al suo interno altri sei Dataverse:

- 3 Dataverse di tipo “ricercatore”;
- 2 Dataverse di tipo “laboratori di ricerca”;
- 1 Dataverse di tipo “gruppo di ricerca”.

A loro volta i sei Dataverse contengono ciascuno almeno un dataset; per esempio, ognuno dei tre Dataverse dei ricercatori include un dataset contenente diversi tipi di file, in particolare file tabulari e file di testo.

Nella pagina di visualizzazione del dataset, oltre alla citazione e ai suoi metadati, è possibile esplorare (ed editare, nel caso si abbia un ruolo che lo consente) i file, i metadati dei file, i tag relativi alla loro tipologia, i loro path, i termini di utilizzo e la versione, dato che il dataset mantiene la storia delle sue modifiche. Tramite il tasto Access dataset o utilizzando la DATA ACCESS API il dataset è scaricabile in formato zip.

Al livello dei file, il software fornisce automaticamente una citazione che può essere esportata in vari formati, così come anche i metadati relativi sono esportabili nei formati più diffusi<sup>27</sup>. Sia per i dataset che per i file è possibile gestire i termini d'uso e l'accesso ai dati.

Mediante la funzionalità Guestbooks il Download users FILELI chiede obbligatoriamente nome, cognome, mail e istituzione a cui la persona è affiliata, mentre lascia opzionale il campo Position; prevede, inoltre, una domanda a risposta multipla, obbligatoria, sugli scopi di utilizzo del materiale che si andrà a scaricare.

Il test diretto delle principali funzionalità di Dataverse ha avuto il principale scopo di sottoporre l'idea dell'adozione del software per un repository dei dati della ricerca all'Università di Pisa, in particolare agli organi deputati al supporto della ricerca, ovvero il Sistema bibliotecario di ateneo<sup>28</sup> e la Direzione Servizi per la ricerca e il trasferimento tecnologico<sup>29</sup>. Il repository Dataverse Unipi sarebbe un archivio istituzionale dei dati della ricerca che, insieme alle piattaforme universitarie

<sup>22</sup> Cfr. <https://github.com/IQSS/dataverse-docker>

<sup>23</sup> Cfr. <https://dans.knaw.nl/en/>

<sup>24</sup> Cfr. <https://www.knaw.nl/en>

<sup>25</sup> Cfr. <https://www.nwo.nl/>

<sup>26</sup> Cfr. <https://www.doi.org/>

<sup>27</sup> Cfr. <https://guides.dataverse.org/en/latest/user/dataset-management.html>

<sup>28</sup> Cfr. <http://www.sba.unipi.it/>

<sup>29</sup> Cfr. <https://www.unipi.it/index.php/amministrazione/itemlist/category/2018-direzione-sprtto>

pisane già esistenti, ovvero all'Archivio della ricerca (ARPI)<sup>30</sup> e all'Archivio digitale delle tesi discusse (ETD)<sup>31</sup>, permetterebbe alla comunità pisana di conservare, preservare e condividere i prodotti derivanti dalle varie fasi del lavoro di ricerca che andrebbero ad aggiungersi alle pubblicazioni depositate in ARPI.

## 6. UNA RETE DI DATA STATIONS

All'interno delle strategie e disposizioni internazionali in merito all'Open science, negli ultimi anni l'attenzione è stata rivolta alle cosiddette *Open Science Infrastructures* (OSInfras) ([8]). Le raccomandazioni UNESCO in merito all'Open science ([13]) individuano queste ultime in tutte le infrastrutture di ricerca condivise, fisiche e virtuali, che sono essenziali per la Scienza aperta, comprendendo nel novero anche gli archivi dei dati della ricerca. Un esempio di infrastruttura di ricerca in ambito europeo è OPERAS<sup>32</sup>, dedicata al supporto della comunicazione accademica aperta nell'ambito delle scienze umane e sociali. Come espressamente indicato nella descrizione del progetto<sup>33</sup>, OPERAS è focalizzata sulla comunicazione e pubblicazione scientifiche nelle suddette discipline e persegue gli obiettivi dell'Open science.

Il centinaio di archivi digitali di dati della ricerca<sup>34</sup> che, a livello internazionale, utilizzano Dataverse, possono essere strutturati in modo diverso: in alcuni casi, il repository è l'espressione unica di una singola istituzione, in altri è progettato come un repository federato con un'unica interfaccia contenente le collezioni dei diversi enti di ricerca - quindi di più istituzioni - sotto forma di Dataverse *collections*. Un'altra soluzione è creare una rete distribuita di repository Dataverse costituendo vere e proprie infrastrutture di ricerca. Per esempio, il centro nazionale di ricerca olandese DANS, che mira alla riusabilità dei dati della ricerca, comprende nella sua nuova infrastruttura:

- le cosiddette *Data stations*, ciascuna riferita a uno specifico ambito disciplinare, ovvero istanze di Dataverse in cui i ricercatori possano gestire i loro dati sotto la supervisione di un Data station manager;
- DataverseNL<sup>35</sup>, una piattaforma aperta di archiviazione dei dati della ricerca che a sua volta contiene Dataverse *collections* e datasets di varie comunità di ricerca.

Similmente, nell'ambito del progetto Social sciences & humanities open cloud (SSHOC)<sup>36</sup>, finanziato da Horizon 2020, è previsto l'utilizzo del software Dataverse per la realizzazione di una rete di *repositories* per le organizzazioni di ricerca dell'area delle scienze sociali e umanistiche di EOSC, come il Consortium of European Social Science Data Archives (CESSDA)<sup>37</sup> e il Common Language Resources and Technology Infrastructure (CLARIN)<sup>38</sup>. DANS partecipa come partner al progetto insieme a CESSDA e ha un ruolo di primo piano nel fornire l'infrastruttura necessaria. Le *Data stations* possono essere implementate sul cloud di EOSC oppure come singole installazioni utilizzando il già citato pacchetto Archive in a box. Il progetto usa il proxy Traefik, che è *cloud-based* e permette un collegamento con le altre realtà Dataverse.

Per raggiungere una maggiore interoperabilità semantica, in SSHOC si sta lavorando a mappature tra i diversi schemi di metadati, al fine di collegare i metadati alle ontologie, rinforzare il supporto all'utilizzo dei vocabolari controllati e rendere ancora più *semantic-friendly* gli strumenti a disposizione. Come sottolineato da Slava Tykhonov, Marion Wittenberg e Wilko Steinhoff, la realizzazione di un sistema distribuito Dataverse assicura il dialogo tra le *Data stations* permettendo a ciascuna, allo stesso tempo, di mantenere le proprie caratteristiche specifiche in base alle esigenze del contesto istituzionale ([12]).

## 7. SCENARI E OPPORTUNITÀ FUTURI

Nel complesso panorama della gestione dei dati della ricerca, rappresentati da oggetti diversi quali parole, numeri, formule, grafici, tabelle, riferimenti bibliografici, software, audio, ecc., diventa sempre più importante per le istituzioni poter contare su strumenti adeguati alla gestione dei dati secondo i principi FAIR di reperibilità, accessibilità, interoperabilità e riutilizzo. Si sostiene che i repository, a causa dell'utilizzo di tecnologie e protocolli progettati oltre due decenni fa, prima dell'esplosione del web, degli strumenti social, del web semantico e dell'uso predominante dei dispositivi mobili per accedere ai contenuti in rete, non abbiano sviluppato del tutto il loro potenziale ([11]); come ricordato in precedenza, tra i

---

<sup>30</sup> Cfr. <https://arpi.unipi.it>

<sup>31</sup> Cfr. <https://etd.adm.unipi.it>

<sup>32</sup> Cfr. <https://operas-eu.org/>

<sup>33</sup> Cfr. <https://operas-eu.org/about/operas-in-a-nutshell/>

<sup>34</sup> Cfr. <https://dataverse.org/installations>

<sup>35</sup> Cfr. <https://dataverse.nl/>

<sup>36</sup> Cfr. <https://sshopencloud.eu/project>

<sup>37</sup> Cfr. <https://www.cessda.eu/>

<sup>38</sup> Cfr. <https://www.clarin.eu/>

motivi di questa mancata occasione si annovera anche l'atteggiamento spesso passivo degli studiosi utilizzatori che non ne hanno colto le potenzialità, forse anche in quanto non sensibilizzati adeguatamente da parte di soggetti e organi deputati. A questo punto, lo sviluppo di piattaforme istituzionali che si avvalgono di software specifici in continuo aggiornamento e gestiti da ampie comunità di utenti potrebbe gettare le basi per la nascita di uno "spazio" in grado di aggregare i dati prodotti dalle comunità scientifiche a livello nazionale o, addirittura, internazionale, così da aprire nuovi scenari per un sistema sempre più "ricerca-centrico", in linea con la visione di COAR: «Our vision is to position repositories as the foundation for a distributed, globally networked infrastructure for scholarly communication, on top of which layers of value added services will be deployed, thereby transforming the system, making it more research-centric, open to and supportive of innovation, while also collectively managed by the scholarly community»<sup>39</sup>.

## BIBLIOGRAFIA

- [1] Bantin, Philip C., a c. di. *Building Trustworthy Digital Repositories: Theory and Implementation*. Rowman & Littlefield, 2016.
- [2] Crosas, Mercè. «Fair principles and beyond: implementation in Dataverse». *Septentrio conference series*, n. 2 (20 marzo 2020). <https://doi.org/10.7557/5.5334>
- [3] Dawey, Lin, Jonathan Crabtree, Ingrid Dillo et al. «The TRUST principles for digital repositories». *Scientific data*, 7 (2020). <https://doi.org/10.1038/s41597-020-0486-7>
- [4] Durand, Gustavo. «Dataverse's Approach to Technical Community Engagement». *Septentrio conference series*, n. 2 (20 marzo 2020). doi: <https://doi.org/10.7557/5.5424>
- [5] Durand, Gustavo. «Integrations and External Tools». Presentato al *Dataverse Community Meeting 2021*, giugno 2021.
- [6] European Commission, Directorate-General for Research and Innovation, Corcho, Eriksson, Kurowski et al. *EOSC interoperability framework : report from the EOSC Executive Board Working Groups FAIR and Architecture*. 2021. <https://doi.org/10.2777/620649>
- [7] European Commission, Directorate-General for Research & Innovation. «H2020 Programme: Guidelines on FAIR data management in Horizon 2020 Version 3.0.» Report. European Commission, Directorate-General for Research & Innovation, 2016. <https://repository.oceanbestpractices.org/handle/11329/1259>
- [8] Grieco, Giuseppe, Ivan Heibi, Arcangelo Massari, Arianna Moretti, e Silvio Peroni. «Enabling portability and reusability of open science infrastructures.» In: *Linking theory and Practice of digital libraries*. TPDL 2022. Lecture notes in computer science, vol 13541. Springer, Cham. [https://doi.org/10.1007/978-3-031-16802-4\\_36](https://doi.org/10.1007/978-3-031-16802-4_36)
- [9] Higgins, Sarah. «The lifecycle of data management». In *Managing research data*, a cura di Graham Pryor, 17–46. Facet, 2012. <https://doi.org/10.29085/9781856048910.003>
- [10] Jain, Priti. «New trends and future applications/directions of institutional repositories in academic institutions». *Library review* 60 (2011): 125–41.
- [11] Rodrigues, Eloy, et al. «Next generation repositories: behaviours and technical recommendations of the COAR Next Generation Repositories Working Group», 28 novembre 2017. <https://doi.org/10.5281/zenodo.1215014>
- [12] Tykhonov, Vyacheslav, Marion Wittenberg e Wilko Steinhoff. «Archive in a box for Dataverse: building distributed data network with shared services». *Dataverse Community Meeting 2022*, 14 giugno 2022, Software session 1, <https://projects.iq.harvard.edu/dcm2022/agenda>, [https://www.youtube.com/watch?v=-Re\\_elg67e8&list=PLLoKvRqQVbtKJFdle-bynhnXdePwI7jpS&index=2&t=4378s](https://www.youtube.com/watch?v=-Re_elg67e8&list=PLLoKvRqQVbtKJFdle-bynhnXdePwI7jpS&index=2&t=4378s).
- [13] UNESCO. *UNESCO Recommendation on Open science. Programme and meeting document SC-PCB-SPP/2021/OS/UROS*, 2021. <https://unesdoc.unesco.org/ark:/48223/pf0000379949>
- [14] Vilela Rodrigues Rezende, Laura, e Sonia Barbosa. «Using the DATAVERSE project to move towards fair principles». In *Princípios FAIR aplicados à gestão de dados de pesquisa*, 2021. <https://www.arca.fiocruz.br/handle/icict/51281>.
- [15] Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. «The FAIR guiding principles for scientific data management and stewardship.» *Scientific data* 3, no. 1 (December 2016): 160018. <https://doi.org/10.1038/sdata.2016.18>

---

<sup>39</sup> Cfr. <https://www.coar-repositories.org/old-pages/activities/advocacy-leadership/working-group-next-generation-repositories/background/>