

Atti del XV Convegno Annuale **AIUCD**

Digitale *e Public Engagement*

Pratiche e prospettive nelle Digital
Humanities

Cagliari 3-4-5 giugno 2026

a cura di

Cristina **Marras** | Andrea **Pergola** | Giampaolo **Salice**

ISBN 9791298618817



Copyright ©2026 AIUCD

Associazione per l'Informatica Umanistica e la Cultura Digitale



Il presente volume e tutti i contributi sono rilasciati sotto licenza Creative Commons Attribution ShareAlike 4.0 International license (CC-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). All other rights retained by the legal owners.

A cura di: Cristina Marras; Andrea Pergola; Giampaolo Salice (2026). *Digitale e Public Engagement: pratiche e prospettive nelle Digital Humanities*, Atti del XV Convegno Annuale AIUCD, Cagliari 3-5 giugno 2026, Università degli Studi di Cagliari.

Ultimo accesso agli URL in data 13 maggio 2026.

Last URL access May, 13 2026.

Si prega di notificare all'editore ogni omissione o errore si riscontri: segreteria [at] aiucd.org

Please notify the publisher of any omissions or errors found: segreteria [at] aiucd.org

Il programma del Convegno AIUCD 2026 è disponibile online:

<https://www.aiucd2026.unica.it/companion/>

The AIUCD 2026 Conference Program is available online:

<https://www.aiucd2026.unica.it/companion/>

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*, effettuata dai membri del Comitato di Programma sotto la supervisione del Comitato Scientifico di AIUCD 2026.

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, carried out by the members of the Programme Committee under the supervision of the Scientific Committee of AIUCD 2026.

Gli Atti del Convegno AIUCD 2026 sono pubblicati come raccolta dei contributi forniti direttamente dagli autori e dalle autrici. I file sono stati raccolti e assemblati senza interventi redazionali significativi da parte dei curatori.

The proceedings of the AIUCD 2026 Conference are published as a collection of contributions provided directly by the authors. The files have been collected and compiled without significant editorial intervention by the editors.

Il logo di AIUCD 2026 è opera di Raffaele Argiolas

The AIUCD 2026 logo was designed by Raffaele Argiolas

La copertina è stata realizzata da Giampaolo Salice

The cover was created by Giampaolo Salice

GENERAL CHAIRS

Giampaolo Salice | Università degli Studi di Cagliari | DH UNICA

Cristina Marras | CNR-ILIESI | AIUCD

COMITATO ORGANIZZATORE / ORGANIZING COMMITTEE

Il Comitato Organizzatore è composto da Giampaolo Salice (Università degli Studi di Cagliari), Cristina Marras (CNR-ILIESI), Raffaele Argiolas, Alessandro Capra, Giommara Carboni, Andrea Pergola, Eleonora Todde (DH UNICA Università degli Studi di Cagliari), Christian D'Agata (Università di Catania), Francesca Frontini (CNR-ILC), Simone Rebora (Università di Verona), Marco Rospocher (Università di Verona), Laura Stochino (ISSASCO), Matteo Tatti e Alice Nozza (Associazione Itzokor).

Iscrizioni / Conference Registration

Eleonora **Todde** – DH UniCa – Università degli Studi di Cagliari

Atti / Conference Proceedings

Andrea **Pergola** – DH UniCa – Università degli Studi di Cagliari

Comunicazione / Communication

Raffaele **Argiolas** – DH UniCa – Università degli Studi di Cagliari

Promozione / Promotion

Giommara **Carboni** – DH UniCa – Università degli Studi di Cagliari

Tecnologo dell'Evento / Event Technologist

Alessandro **Capra** – DH UniCa

Organizzazione locale

Dipartimento di Lettere, Lingue e Beni Culturali, Università degli Studi di Cagliari, Ufficio Amministrativo:
Serena Serra, Giulia Cadoni, Claudia Serri

COMITATO SCIENTIFICO / SCIENTIFIC COMMITTEE

Il Comitato Scientifico del convegno AIUCD 2026 è composto dai Chairs di Programma e dai Chairs delle cinque aree tematiche.

Program Chairs

Giampaolo Salice | Università degli Studi di Cagliari | DH UNICA

Cristina Marras | CNR-ILIESI | AIUCD

DH e co-costruzione del sapere con le comunità: sfide, metodi, strumenti / *DH and co-construction of knowledge with communities: challenges, methods, tools*

Arianna Ciula | King's Digital Lab, London

Greta Franzini | EURAC, Bolzano

Archivi ed edizioni: descrizioni aumentate, accessibilità e sistemi informativi / *Archives and editions: augmented descriptions, accessibility, and information systems*

Federico Valacchi | Università di Macerata

Paolo Monella | Università Kore Enna | AIUCD

Testualità digitali: prospettive, sviluppi e sperimentazioni / *Digital textualities: perspectives, developments, and experimentations*

Simone Rebora | Università di Verona | AIUCD

Alessandro Adamou | Biblioteca Hertziana

Rappresentazione di Dati e Conoscenza / *Data and Knowledge Representation*

Francesca Tomasi | Alma Mater Studiorum – Università di Bologna

Angelo Mario Del Grosso | CNR-ILC

Memorie, storia e patrimoni culturali digitali / *Tangible and Intangible Digital Heritage*

Michela Tardella | CNR-ILIESI

Enrica Salvatori | Università di Pisa

COMITATO DI PROGRAMMA / PROGRAM COMMITTEE

Stefano Allegrezza (Università di Macerata), Laura Antonietti (Université de Versailles Saint Quentin en Yvelines - Université Paris-Saclay), Alessio Antonini (Open University), Liborio P. Barbarino (Università di Catania), Nicola Barbuti (Università di Bari "Aldo Moro"), Stefano Bazzaco (Università di Verona), Andrea Bellandi (CNR-ILC), Giulia Benotto (CNR-ILC), Monica Berti (Leipzig University), Mario A. Bochicchio (Università di Bari "Aldo Moro"), Andrea Bolioli (Ricercatore indipendente), Marco Bombieri (Università di Verona), Flavia Bruni (Università di Chieti-Pescara), Marina Buzzoni (Università Ca' Foscari di Venezia), Alberto Campagnolo (KU Leuven), Vittore Casarosa (CNR-ILC), Raffaele Cioffi (Università di Napoli Federico II), Vincenzo Colaprice (Università di Torino), Francesca Congiu (Università degli Studi di Cagliari), Giuseppe Consolo (Università degli studi di Napoli, Federico II), Christian D'Agata (Università di Catania), Giulia D'Agostino (TU Darmstadt), Elisa D'Argenio (HUN-REN Hungarian Research Centre for Linguistics), Davide Dainese (Alma Mater Studiorum Università di Bologna), Stefano Dall'Aglio (Università Ca' Foscari Venezia), Marilena Daquino (Alma Mater Studiorum Università di Bologna), Mauro De Bari (Università di Bari "Aldo Moro"), Angelo M. Del Grosso (CNR-ILC), Matteo Di Cristofaro (UniMore), Francesca Di Donato (CNR-ILC), Giorgia Di Marcantonio (Università di Macerata), Giorgio Maria Di Nunzio (Università di Padova), Roberto Evangelista (CNR-ISPF), Stefano Ferilli (Università di Bari "Aldo Moro"), Franz Fischer (Università Ca' Foscari di Venezia), Greta H. Franzini (EURAC), Francesca Frontini (CNR-ILC), Daniele Fusi (Università Ca' Foscari di Venezia), Mariangela Giglio (Alma Mater Studiorum Università di Bologna), Tiago Luis Gil (University of Brasilia), Michela Giordano (Università degli Studi di Cagliari), Luca Giovannini (University of Potsdam), Milena Giuffrida (Università di Catania), Giovanna Granata (Università degli Studi di Cagliari), Edmondo Grassi (Università Telematica San Raffaele Roma), Miryam Grasso (Università di Catania), Piergiovanna Grossi (Università di Verona), Marina M. Guglielmi (Università degli Studi di Cagliari), Alessandro Iannella (Università degli Studi di Milano), Sabrina Iorio (Università Napoli Federico II), Michele Lacriola (Università di Siena), Alessandro Laruffa (ISSTOR), Anna Mambelli (UniMore), Francesco Mambrini (Università Cattolica del Sacro Cuore), Francesco Mamei (Università degli Studi di Cagliari), Lorenzo Mancini (CNR-ILIESI), Anna Maria Marras (Università di Torino), Cristina Marras (CNR-ILIESI), Pietro Mazzarisi (Università di Trieste), Barbara McGillivray (King's College London), Federico V. Meschini (Università della Tuscia), Alessio Miaschi (CNR-ILC), Giulia Miglietta (Università del Salento), Paolo Monella (Università Kore di Enna), Rossana Morriello (Università degli Studi di Firenze), Giulia Murgia (Università degli Studi di Cagliari), Enrico Natale (infoclio.ch), Sebastiana Nocco (CNR-ISEM), Serge Noiret (AIPH), Giuseppe Palazzolo (Università di Catania), Mafalda Papini (CNR-ILC), Enrico Pasini (Università di Torino), Giulia Pedonese (CNR-ILC), Paola Peratello (Università Ca' Foscari Venezia), Andrea Pergola (Università degli Studi di Cagliari), Ginevra Peruginelli (CNR-IGSG) University of Groningen, Federico Pianzola (University of Groningen), Fabio C. Pinna (Università degli Studi di Cagliari), Igor Pizzirusso (AIPH), Tiziana Pontillo (Università degli Studi di Cagliari), Valeria Quochi (CNR-ILC), Lisa Reggiani (CNR-ILIESI), Giulia Renda (Alma Mater Studiorum Università di Bologna), Pietro Restaneo (CNR-ILIESI), Dario Rodighiero (University of Groningen), Roberto Rosselli Del Turco (Università di Torino), Federica Rovelli (Università di Pavia), Giampaolo Salice (Università degli Studi di Cagliari), Enrica Salvatori (Università di Pisa), Emilio M. Sanfilippo (CNR-ISTC), Eva Sassolini (CNR-ILC), Manfredi Scanagatta (UniMore), Andrea Schimmenti (Alma Mater Studiorum Università di Bologna), Flavia Sciolette (CNR-ILC), Alessia Scognamiglio (CNR-ISPF), Luigi Serra (CNR-ISEM), Pietro Sichera (CNR-ILIESI), Daria Spampinato (CNR-ISTC), Giulia Speranza (Università di Napoli "L'Orientale"), Rachele Sprugnoli (Università di Parma), Francesco V. Stella (Università di Siena), Timothy Tambassi (Università Ca' Foscari di Venezia), Mirko Tavosanis (Università di Pisa), Francesca Tomasi (Alma Mater Studiorum Università di Bologna), Simona Turbanti (Università di Milano), Nicoletta Usai (Università degli Studi di Cagliari), Marco Venuti (Università di Catania), Gennaro Vessio (Università di Bari "Aldo Moro"), Gabriele Vezzani (Università di Verona), Simone Zenzaro (CNR-ILC).

Enti organizzatori / Organisers

Il XV convegno annuale dell'Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD) è organizzata dal DH UNICA - Centro Interdipartimentale per l'Umanistica Digitale e dal Dipartimento di Lettere, Lingue e Beni Culturali dell'Università di Cagliari in collaborazione con AIUCD.

Sommario

PREMESSA

I-IV

Cristina Marras, Andrea Pergola, Giampaolo Salice

[1/A] INTELLIGENZA ARTIFICIALE GENERATIVA E RIFLESSIONE TEORICA

- Una Agenda sull'AI generativa per Digital Humanities 6
Fabio Ciotti
- L'errore dell'IA come euristica per una deliberazione assistita 14
Valerio De Luce; Tiberio Uricchio
- Verso una memoria artificiale? Il progetto Johannes tra questioni tecnologiche, archeologiche e morali 21
Stefano Bertoldi
- AI confini della realtà: il pensiero umano nell'epoca della sua riproducibilità generativa 31
Daniele Silvi

[1/B] CORPORA DIGITALI: REQUISITI E SOLUZIONI GESTIONALI

- Verba volant, emails manent*: una proposta di metodologia per l'acquisizione e la preservazione degli archivi di posta elettronica di personaggi illustri 38
Alberto Abis, Stefano Allegrezza, Fabrizio Stupino
- Sfide alla sostenibilità dei progetti di digitalizzazione: i casi studio degli archivi manicomiali calabresi 45
Emanuela Nicole Donato, Maria Natalia Federico, Grazia Serratore
- A technical solution for a digital cultural heritage corpus: Notes on collaboration, requirements and sustainability 53
Arianna Ciula, Miguel Vieira, Geoffroy Noel, Neil Jakeman, Zihao Lu, Tiffany Ong, Simona Stoyanova, Jonathan Prag, Alessia Coccato, Ryan Heuser

[1/C] CO-PRODUZIONE DI ARCHIVI E COLLEZIONI DI COMUNITÀ

- L'Archivio Digitale di Vigne Nuove. Costruire un archivio di comunità, tra teoria e pratica 62
Manfredi Scanagatta
- Participatory Digital Archives as a Method for Co-constructing Cultural Memory: The Bunjevke Women's Platform 68
Sandra Iršević PhD
- Il ritorno dei partigiani sardi. Una collezione digitale co-prodotta 73
Walter Falgio, Laura Stochino, Matteo Quarantiello
- Dalla materialità delle fonti al dato condiviso: metodi ibridi per un archivio digitale partecipativo 78
Tiziana Pasciuto

[2/A] LARGE LANGUAGE MODELS: APPLICAZIONI UMANISTICHE

| | |
|--|-----|
| Scholarly Opinion Mining using LLMs and Knowledge Graphs: the case of the Van den vos Reynaerde | 87 |
| Andrea Schimmenti, Joris van Zundert, Fabio Vitali, Marieke van Erp | |
| Un LLM per le Scienze Umane e Sociali? ReSearch_SSH come esperimento europeo | 95 |
| Adam Faci, Alessio Miaschi, Anne Combe, Pascal Cuxac ⁴ , Francesca Frontini, Nicolas Larrousse, Stéphane Pouyllau | |
| L'AI a servizio della Musicologia: costruzione di thesauri musicologici con il supporto dei Large Language Models | 102 |
| Manuela Grillo, Paolo Bonora | |
| Per una formazione del filologo computazionale del futuro: Domain Specific Language e Large Language Model | 107 |
| Christian D'Agata, Angelo Mario Del Grosso, Federico Boschetti | |

[2/B] OPEN SCIENCE, SOSTENIBILITÀ E PRATICHE APERTE

| | |
|--|-----|
| Are Digital Humanities really committed to open? An exploratory study on the availability of methodological workflows and open peer review practices | 116 |
| Silvio Peroni | |
| Monitorare la transizione: l'Osservatorio sulle pratiche di Open Science nelle SSH come strumento di Public Engagement | 122 |
| Pamela Barletta, Marta Caradonna, Nicola Giampietro, Alice Orrù, Federico Silvestri | |
| Nel 25° anniversario di Wikipedia: l'ecosistema Wikimedia e le Digital Humanities tra pratiche collaborative e sostenibilità della conoscenza | 130 |
| Piergiovanna Grossi | |
| Una pausa e tre domande. La crisi della memoria nell'era dell'accesso | 137 |
| Enrica Salvatori, Federico Valacchi | |

[2/C] MODELLI DIGITALI PER GESTIRE E CONDIVIDERE FONTI MULTIDISCIPLINARI

| | |
|--|-----|
| Digital Repatriation in Practice: Sharing Ethnographic Photography with Communities of Origin at the Náprstek Museum | 142 |
| Daniela Šnapková | |
| Una storia da ascoltare. Descrivere, mappare, connettere: un modello per la valorizzazione degli archivi musicali | 147 |
| Martina Gremignai, Luca Andrea Ludovico | |
| LETSDigit: Archivio digitale della letteratura a Trieste | 154 |
| Alina Jill Simeone, Cristina Fenu, Marina Buzzoni, Roberto Rosselli Del Turco | |
| La mostra <i>Coerenza In Coerenza</i> (1984) oltre la fotogrammetria: ricostruzione digitale tramite reti neurali | 161 |
| Filippo Yahia Masri, Maria Carmelita | |

[3/A] AI E NLP PER TESTI E RECUPERO DELL'INFORMAZIONE

| | |
|--|-----|
| Creating a new gender-fair Italian morphological classifier based on UmBERTo fine-tuning | 168 |
| Irene Caiazzo, Giovanna Maria Dimitri | |
| Assessing LLM Capabilities in Cultural Heritage: Recent Trends and Gaps in Benchmarks | 175 |

Remo Grillo, Gianmarco Spinaci, Lukas Klic, Giovanni Colavizza

L'intelligenza artificiale per il potenziamento dell'information retrieval nei discovery tool delle biblioteche 183

Ilaria Belvedere, Simona Turbanti

Sinergie ludonarrative: l'evoluzione del racconto videoludico tra divulgazione, tecnica ed intelligenza artificiale 191

Giulia Angelini, Tommaso Mazzoli

[3/B] DH PER STORIA, DIRITTO E ARTE

Archivi ibridi di persona e preservazione forense dei materiali nativi digitali 198

Mariangela Giglio, Adele Gorini

Digital Humanities e studi neogreci: il progetto della Biblioteca Digitale dell'Osservatorio "Mario Vitti" 204

Georgios Katsantonis

artresearch.net: A Research Infrastructure for Exploring Interpretive Plurality in Art-Historical Photo Archives 211

Marilena Daquino, Francesca Mambelli, Alessandro Adamou, Pietro Maria Liuzzo, Stefanie Schneider, Spyros Koulouris, Artem Kozlov, Rafael Brundo Uriarte

Legislative Changes Concerning the Reproduction of Images of Italian Cultural Heritage Assets between 2023 and 2025 217

Marco Ciurcina, Piergiovanna Grossi

[3/C] ANALISI COMPUTAZIONALE DEL TESTO LETTERARIO

Flaiano postcoloniale? Un approccio computazionale alla *vexata quaestio* di *Tempo di uccidere* 226

Silvia Lilli, Daniel Raffini

Verso una rappresentazione del macrotesto: filologia d'autore e analisi quantitativa nei racconti di Fausta Cialente 234

Emmanuela Carbè

Analisi fluidodinamica dei campi semantici in testi letterari: il Numero di Reynolds come modello di turbolenza testuale 241

Pietro Sichera, Carla Perrone, Antonio Sichera

Dentro l'Officina del Testo: Modellare Strutture Letterarie Complesse 248

Enrica Bruno

[4/A] STUDI LETTERARI DIGITALI E DISTANT READING

Tracing Intertextuality at Scale. A Network Study of Poetic Influence 255

Gabriele Vezzani

The Landscape of Italianistica: Italian Literary Studies and Digital Humanities in Semantic Space 261

Maria Levchenko

L'Edicola della Storia di Trieste: NLP e *distant reading* Per il confronto tra periodici storici italiani e sloveni nella Trieste del 1902 267

Cristina Fenu, Giovanni Pinna

Literary Tropes as Prescriptive Devices: A Case Study from BookTok 274

[4/B] MUSEI E ACCESSO DIGITALE

- Rigenerare un patrimonio desueto con il phigital: la Casa Museo Bartolo Longo 282
Mauro De Bari
- Musei senza confini: il digitale come infrastruttura di memoria partecipata nei musei delle aree interne. Il GeoMuseo Monte Arci Stefano Incani di Masullas e il Museo Paleontologico PARC di Genoni 290
Paola Palmas
- Pubblici remoti: la Realtà Virtuale come infrastruttura culturale di accesso 296
Elisa Smeraldo

[4/C] FORMAZIONE, DIDATTICA E LABORATORI DH

- Gli Hunger Games del LabCD. Un diario di sopravvivenza 304
Enrica Salvatori, Vittore Casarosa, Francesco Ricciardi
- It's Buildin' Time: A Project-based Approach for Teaching Data Management to Cultural Heritage Graduate Students 309
Sebastian Barzagli, Alba d'Elia, Esther Montserrat Giordano, Giulia Guidarelli, Gianluca Petrosillo, Elisabetta Sabattini, Fiammetta Sabba, Lucia Sardo
- Integrazione di tecnologie per la traduzione e CLIL nella creazione di risorse didattiche per il sardo 319
Gianfranco Fronteddu, Igor Deiana

[5/A] LINKED OPEN DATA, MODELLI SEMANTICI, INFRASTRUTTURE

- Reti sociali e linked data: persone e co-occorrenze documentarie nel Portale delle fonti per la storia della Repubblica italiana 329
Herbert Natta, Gianluca Rossi, Roberta Maggi
- Web Portals on Linked Open Data: Testing Sampo-UI in Projects on Early Modern Bibliographic Data and Chinese Calligraphy and Arts 336
Arianna Moretti, Katarina Lučić, Iiro Lassi Ilmari Tiihonen, Jonas Paul Fischer
- Risorse Terminologiche e Linked Open Data: il *Glossario delle Infrastrutture di Ricerca (GIR)* come Caso di Studio 345
Lucia Francalanci, Alessia Scognamiglio, Giulia Pedonese, Michele Mallia, Irene Falini, Fahad Khan
- Orchestrare servizi e scenari avanzati: fruizione e necessità delle comunità di ricerca nel marketplace di H2IOSC 352
Pietro Sichera, Cristina Marras, Enrico Pasini, Vittoria Fabiani, Paolo Ongaro, Michele Scapicchi, Chiara Di Pietro

[5/B] PATRIMONIO CULTURALE: SISTEMI INFORMATIVI INTEGRATI

- Architettura e *Digital Humanities*. Conoscenza e valorizzazione del Patrimonio attraverso la costruzione di sistemi informativi integrati 361
Donatella Rita Fiorino, Caterina Giannattasio, Elisa Pilia, Valentina Pintus, Andrea Pirinu, Marcello Schirru
- Tra testo e architettura digitali per lo studio delle trasformazioni storico-artistiche dei monasteri benedettini 368

| | |
|--|-----|
| Gianmario Guidarelli, Paolo Borin, Sonia Cavicchioli, Chantal Pivetta, Renato Canearo, | |
| Dalla pergamena al virtuale. Manoscritti miniati nello spazio tridimensionale | 375 |
| Valeria Minisini, Bruno Fanini, Giorgio Gosti | |
| Modeling and visualizing palimpsests with IIIF Presentation API 3.0 | 382 |
| Panagiotis Leontaridis, Giacomo Marchioro, Glen Robson | |

[5/C] CROWDSOURCING E PUBLIC ENGAGEMENT

| | |
|---|-----|
| <i>Tag ANIMALx</i> : Crowdsourcing the Non-Human in Cultural Heritage Collections | 389 |
| Daniela Teixeira Gomes, Carla Vieira, Joana Vieira Paulino | |
| From perception to memory: Public Engagement and Neurohumanistic Approaches to the Collective Reconstruction of Archaeological Heritage | 396 |
| Grazia Solenne, Vincenza Ferrara, Elisa Corrà | |
| Comunicare la ricerca archeologica nell'ecosistema digitale: pratiche, metodi e sperimentazioni su Instagram | 404 |
| Marco Demuru | |
| Public engagement nell'era dei synthetic media: per una co-interpretazione critica tra Digital Humanities e Media Education | 411 |
| Gabriele Prosperi, Mario A. Bochicchio | |

[6/A] MODELLAZIONE TESTUALE E LINGUISTICA COMPUTAZIONALE

| | |
|--|-----|
| Verso un dizionario narrativo computazionale degli usi linguistici in aree ad alta vulnerabilità sociale | 420 |
| Michela Bandini, Silvia Piccini, Andrea Bellandi, Emiliano Giovannetti | |
| Verso la creazione di una Treebank per il sardo | 426 |
| Nicoletta Puddu, Manuela Sanguinetti, Luigi Talamo | |
| A Learning-by-Doing Approach to Spoken Data Collection: The Case of the LPSP Corpus | 432 |
| Claudia Roberta Combei, Gaia Eleonora Di Raimondo, Sofia Maestri, Benedetta Romanazzi, Elena Scotti | |
| Digital Humanities e Letteratura Elettronica in Italia: marginalità e convergenze | 440 |
| Giuseppe Arena | |

[6/B] CO-CREAZIONE, RICERCA PARTECIPATIVA, COMUNITÀ

| | |
|---|-----|
| Metodi e strumenti di sperimentazione tra co-creazione, Digital Humanities, e ricerca demo-etno-antropologica: la digitalizzazione come pratica relazionale ed etica. | 448 |
| Eleonora De Longis, Matteo Cova, Giovanni Bertelli | |
| Integrare strumenti digitali e conoscenze della comunità per una progettazione urbana resiliente al clima: l'iniziativa Dundrum by Design | 455 |
| Chiara Cocco, Mattia Leone, Antonio Savino, Miruna Popa, Sara Tedesco | |
| HCI e AI per il Public Engagement: casi di studio per la valorizzazione del patrimonio culturale | 462 |
| Samuel Aldo Iacolina, Manuela Angioni, Valentina Marotto, Francesca Mura, Piergiorgio Palla | |

| | |
|---|-----|
| The Automation of Research in Digital Humanities: Artificial Researchers, Digital Archives, and the Democratization of Scholarly Access | 468 |
| Roberto Di Quirico | |

[6/C] AI E AUTOMAZIONE PER ARCHIVI E PATRIMONIO

| | |
|--|-----|
| Dal problema storico alle IA: Il Database on the Slave Trade between the Mediterranean and the Atlantic (15 th -16 th Centuries) | 477 |
| Salvatore Spina, Carlo Taviani, Jörg Hörnschemeyer | |
| Peirce Interprets Peirce: Digitization, Automation, and Interpretation in Charles Peirce's Manuscripts | 485 |
| Alessandro Adamou, Sebastian Feil, Davide Picca, Carlo Teo Pedretti, Dario Rodighiero, | |
| Lorenzo Zangari | |
| A New Semantic Model for Mythologiae: from Data to Interpretation | 494 |
| Bianca La Manna | |

[7/A] ONTOLOGIE, FAIR DATA E MODELLAZIONE DELLA CONOSCENZA

| | |
|--|-----|
| The ATLAS Guidelines for producing FAIR research products in the Digital Humanities | 502 |
| Chiara Martignano, Giorgia Rubin, Sebastiano Giacomini, Alessia Bardi, Marina Buzzoni, Marilena Daquino, Riccardo Del Gratta, Angelo Mario Del Grosso, Franz Fischer, Roberto Rosselli Del Turco, Francesca Tomasi | |
| RATIO: strumenti per la creazione e la gestione di risorse e collezioni FAIR nell'ambito delle scienze umane | 509 |
| Federico Silverstri, Lorenzo Mancini, Laura Baggiani, Marco Bagiacchi | |
| Making Uncertainty Explicit: The Preservation of Interpretive Complexity across Digital Humanities Project Workflows | 516 |
| Tommaso Battisti, Valentina Pasqual, Giulia Renda, Marilena Daquino | |
| Iconclass as a Semantic Framework for the Analysis of the Sacred Space | 524 |
| Polina Voronova | |

[7/B] GIS E SPAZIALIZZAZIONI DIGITALI

| | |
|--|-----|
| Gli <i>Historical GIS</i> e la storia forestale: potenzialità e applicazioni | 531 |
| Vincenzo Colaprice | |
| Le nazioni genovesi nell'Atlante digitale di Storia Marittima del Regno di Sardegna (ASMSA). Lavori in corso | 539 |
| Giampaolo Salice, Giommara Carboni | |
| Integrating Romanian Toponymy in a GIS Background: dictionary forest names vs. OpenStreetMap forest data | 547 |
| Roxana Patras, Ana Odochiciuc, Mihai-Bogdan Atanasiu, Constantin Răchită | |
| Digital Humanities per narrare i territori: un WebGIS sui porti minori della Sardegna dal medioevo alla contemporaneità. Una proposta di implementazione collaborativa | 554 |
| Luigi Serra, Sebastiana Nocco | |

[7/C] MODELLI, METODI E PROSPETTIVE DELL'EDIZIONE DIGITALE

| | |
|---|-----|
| Roman and Justinianic Legal Terminology in Thirteenth-Century Western European Diplomatic Sources | 562 |
|---|-----|

| | |
|---|-----|
| Tamás Kovács, Angelos Nicolaou, Johannes Laroche, Georg Vogeler | |
| In Search of Lost Time: an Innovative Model for Representing Autograph Texts | 567 |
| Giulia Baldelli, Daniele Fusi, Matteo Zupancic, Franz Fischer, Claus Zittel | |
| <i>Naples Dante Project</i> . Un ambiente integrato per l'analisi bibliografica, filologica e iconografica | 573 |
| Gennaro Ferrante, Andrea La Veglia, Daniele Fusi, Stefano Angelo Rizzo, Angelo Eugenio Mecca, Daniele Duranti, Anna Sviridova | |
| La curatela biblioteconomica delle raccolte digitali MAB: linee gestionali, requisiti, workflow | 580 |
| Nicola Barbuti | |

[8/A] GRAFI, LINKED DATA E MAPPING

| | |
|---|-----|
| DOG: an open-source ecosystem for mapping, narrating, and exploring humanities data | 588 |
| Alessandro Laruffa, Alessandro Capra | |
| Beyond Spatial Data: Exploring Gender and Property in the Catasto Generale Toscano through the Florentia Illustrata Semantic Platform | 596 |
| Erica Andreose , Remo Grillo , Gianmarco Spinaci | |
| Per la storia del mercato dell'arte tra otto e novecento: raccolta, modellazione e valorizzazione dei dati sugli antiquari italiani | 602 |
| Valentina Rossetti, Marilena Daquino, Francesca Mambelli, Ludovica Pannitto, Francesca Tomasi | |
| Un grafo per la ricerca storico-filologica in archivio: il caso del fondo Giuseppe Albini | 609 |
| Lucia Giagnolini, Mohamed Iheb Ouerghi, Valentina Pasqual, Cecilia Tamagnini, Francesca Tomasi | |

[8/B] GEOGRAFIE LETTERARIE E MULTIMEDIALI

| | |
|---|-----|
| 'Spazializzare' la narrativa ecocritica italiana: la piattaforma dell'Atlante digitale della letteratura ecologica italiana | 617 |
| Valentina Pasqual | |
| MeMo: mappe digitali, memoria letteraria e public engagement nel Mezzogiorno | 624 |
| Laura Giurdanella; Giuseppe Palazzolo | |
| Una proposta di approccio computazionale alle geografie del cinema italiano | 631 |
| Alberto Savi, Luca Giovannini | |
| Il campo semantico del fuoco nella Commedia: un caso di studio sulla pertinenza lessicale e sulla visualizzazione | 638 |
| Ruoci Song | |

[8/C] ANNOTAZIONI E MARKUP

| | |
|--|-----|
| Ritagli di teatro. Annotazione semantica della rassegna stampa del Teatro Sant'Erasmus | 644 |
| Paolo Bonora | |
| Annotazione semantica delle formule poetiche del Medioevo germanico | 651 |
| Roberto Rosselli Del Turco, Letizia Vezzosi | |
| Ça y est : Annotating Arabic Texts for Teaching | 658 |

| | |
|---|-----|
| Dal markup all'interpretazione: interrogare il linguaggio della critica verista in un corpus XML/TEI | 665 |
| Denise Bruno, Salvatore Cristofaro, Antonio Di Silvestro, Laura Mazzagufò, Daria Spampinato, Giuseppe Zappalà | |

[9] SESSIONE POSTER

| | |
|--|-----|
| Archivi letterari nativi digitali: modelli descrittivi ed edizioni sperimentali | 673 |
| Elena Barchielli, Simon Willemin, Elena Spadini. | |
| Lexicon Philosophicum: una rivista digitale per la condivisione della conoscenza umanistica | 679 |
| Pamela Barletta, Maria Cristina Dalfino, Pietro Restaneo | |
| Intelligenza artificiale, reference digitale e accessibilità: l'esperienza di Alphabetic | 684 |
| Flavia Bruni – Elisabetta Castro | |
| Grafoteca e la costruzione digitale della memoria manoscritta: curatela, metadattazione e reti di risorse | 691 |
| Amalia Carrano | |
| Valorizzazione digitale del patrimonio culturale: Public-Private Partnership e partecipazione sociale | 697 |
| Annalisa Ciliberti | |
| I Vangeli digitali e accessibili, per un'editoria religiosa universale e senza barriere | 703 |
| Giulia D'Arcangelo | |
| Implementing a controlled vocabulary for medieval administrative acts: theoretical and methodological considerations | 709 |
| Chiara De Bastiani | |
| L'Archivio del Centro Conservazione e Restauro "La Venaria Reale": un caso studio per la diffusione dei dati tecnico-scientifici | 717 |
| Stefania De Blasi, Lorena Palmieri, Chiara Pipino | |
| Strumenti digitali per i percorsi di visita della Palazzina di Caccia di Stupinigi: Residenza Sabauda e Museo dell'Ammobiliamento | 722 |
| Stefania De Blasi | |
| Strategie di protezione, governance e sostenibilità degli archivi orali | 727 |
| Rosaria De Luca, Elvira Mercatanti, Monica Monachini | |
| FAIR Memories: A Workflow for the Preservation and Collaborative Reuse of Analog Oral History Collections | 734 |
| Nike del Quercio, Costanza Paolillo, Laurent Fintoni | |
| Ontologycore: A Formal Knowledge Graph of Internet Aesthetics | 740 |
| Anouk Flinkert, Ekaterina Krasnova, Shiho Nakamura | |
| Il ruolo del Test Plan per la validazione: il caso d'uso del Marketplace H2IOSC | 747 |
| Vittoria Fabiani | |
| Partecipazione, mediazione e conoscenza: Àndalas de Cultura quale ecosistema digitale del patrimonio culturale della Sardegna | 754 |
| Carolina Floris | |
| From Violin Lessons to Linked Open Data. Reconstructing Tartini's <i>Scuola delle Nazioni</i> through Digital Prosopography and Digital Editions | 762 |
| Selina Galka, Marcella Tambuscio, Rolf Wissmann, Cristina Scuderi, Georg Vogeler | |

| | |
|--|-----|
| Addestramento di un modello di HTR in Transkribus per testi plurilingui di età moderna: il caso del <i>Ripulimento della lingua sarda</i> (Madau) | 768 |
| Michela Incollu, Giulia Murgia | |
| Modelli digitali interattivi per gli archivi teatrali storici: mappare le relazioni semantiche al Teatro Metastasio di Prato (1827–1860) | 775 |
| Matilde Innocenti | |
| Tra fotografia e segno: la trascrizione informatizzata dei registri storici Alinari | 779 |
| Pamela Krzemien | |
| Biblioteche filosofiche private e memoria culturale: http://picus.unica.it/ | 786 |
| Andrea Lamberti, Giovanna Granata | |
| Beyond Archives: A Three-Stage Workflow for Relational Digital Libraries of Written and Oral Memory Sources | 791 |
| Giulia Lembo | |
| A Workflow-Centred Approach to Managing Semantic Artifacts as Linked Open Data | 800 |
| Michele Mallia, Fahad Khan, Valeria Quochi | |
| SEBASTIAN: uno strumento per l'organizzazione, l'analisi spaziale e la visualizzazione dei dati storici | 807 |
| Christian Marcantonio, Giuseppe Consolo | |
| Raccontando e valorizzando il patrimonio culturale. Dalle prime esperienze ai progetti futuri del CNR-ISEM | 814 |
| Maria Grazia Mele R., Samuel Aldo Iacolina, Luigi Serra, Giovanni Serreli | |
| La Digital Library e l'Ecosistema digitale per la cultura come opportunità di <i>public engagement</i> | 821 |
| Federico Meschini, Biancamaria Hermanin, Antonella Negri, Giuliano Romalli | |
| Strategie di engagement nella creazione e sostenibilità di un parco letterario: i risultati del progetto CHANGES | 827 |
| Sara Obbiso, Nicola Mariniello, Davide Bagnaresi, Riccardo Stracuzzi, Giuliana Benvenuti, Alessandro Iannucci | |
| Lo studio integrato del patrimonio materiale e immateriale per la ricostruzione delle pratiche storiche di gestione delle risorse ambientali nell'Alpujarra (Sierra Nevada): il progetto Sinergy | 835 |
| Alessandro Panetta | |
| <i>Reveduti, Correcti, Aprobati et Confirmati</i> : The Digital Edition of the Statutes of Ascoli | 841 |
| Michela Parma, Fabio Mariani | |
| Trascrivere automaticamente i registri catastali sardi del primo Novecento con Transkribus: risultati di un caso di studio | 849 |
| Andrea Pergola, Cecilia Tasca | |
| Ricostruzione virtuale di un patrimonio disperso: le collezioni di disegni a Genova nel XIX secolo | 856 |
| Giulia Pilosu | |
| Dai workflow descrittivi ai workflow applicativi tramite AEON e l'IA | 860 |
| Francesco Pinna, Emiliano Degl'Innocenti | |
| <i>Gamification</i> per la valorizzazione dei beni musicali. Un'ipotesi di raccolta di dati catalografici in <i>crowdsourcing</i> | 867 |

| | |
|--|-----|
| Marcello Ranieri | |
| Tra riconoscimento disciplinare e prospettiva pubblica: un modello logico-applicativo per le conoscenze e le competenze nelle DH | 874 |
| Lisa Reggiani | |
| Mappe, scale e interpretazione. Per una cartografia digitale semantica dello spazio narrativo | 883 |
| Lorenzo Sabatino | |
| Verso un grande corpus diacronico dell'italiano: il contributo degli esempi del GDLI | 893 |
| Eva Sassolini, Sebastiana Cucurullo, Marco Biffi e Simonetta Montemagni | |
| Verso una classificazione funzionale dei prompt basata sulle intenzioni cognitive esplicitate dagli studenti nell'interazione con l'IA | 901 |
| Daniele Scala, Salvatore Varriale | |
| Il videogioco come strumento di mediazione interculturale | 909 |
| Alessia Stocco | |
| Edizione digitale commentata di <i>Le libere donne di Magliano</i> di Mario Tobino: primi passi, metodi e prospettive | 917 |
| Fabio Zarroli | |
| Il restauro testuale dei papiri nel progetto GreekSchools: CoPhiEditor e l'integrazione di modelli linguistici | 923 |

Premessa

AIUCD 2026, il XV Convegno annuale dell'Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD), è dedicato al tema "Digitale e Public Engagement: pratiche e prospettive nelle Digital Humanities". Il convegno mette al centro pratiche, strumenti e metodologie che rendono la conoscenza più aperta, condivisa e inclusiva e contribuiscono alla diffusione di una cultura digitale, intesa come componente costitutiva della cittadinanza contemporanea

Il tema è al centro delle attività del DH UNICA – Centro interdipartimentale per l'Umanistica Digitale dell'Università di Cagliari, organizzatore del Convegno in collaborazione con AIUCD. Il Centro, istituito nel 2020, conta attualmente 76 ricercatrici e ricercatori, espressione di 5 dipartimenti dell'Università degli studi di Cagliari sia di Scienze Umane e Sociali sia di STEM (Letlib, Scienze Politiche, Dicaar, Pedagogia, Psicologia, Filosofia, Ingegneria elettrica ed elettronica).

Il DH UNICA è dotato di un ecosistema digitale, che ottimizza la gestione di una ventina di progetti di ricerca, ed è articolato in ambiti di ricerca e in gruppi di lavoro. Per il carattere fortemente multidisciplinare del Centro e per le tematiche che vi gravitano, come storia digitale pubblica, archivi, musica, quest'anno il XV Convegno dell'AIUCD 2026 si è orientato a dare un contributo ad un dibattito attuale e molto vivo nella comunità DH anche internazionale: l'impatto e il coinvolgimento del pubblico.

Le Digital Humanities assumono oggi, infatti, un ruolo sempre più centrale nella costruzione di pratiche di ricerca aperte e partecipative. La crescente disponibilità di strumenti digitali avanzati, dai database collaborativi alle piattaforme di intelligenza artificiale, offre opportunità senza precedenti per coinvolgere pubblici eterogenei nella produzione e interpretazione della conoscenza. In questo scenario, il public engagement non è più un'attività marginale o secondaria, ma un elemento strutturale della ricerca, in grado di generare nuove forme di collaborazione, arricchire la qualità dei dati e favorire una diffusione più ampia e inclusiva dei risultati scientifici. La riflessione sulle modalità di partecipazione, sulle tecnologie utilizzate, sulle politiche della ricerca e sugli effetti sociali e culturali di questi processi rappresenta quindi un tema cruciale per comprendere e orientare l'evoluzione delle Digital Humanities nel presente e nel futuro.

Il concetto di comunità scientifica è uno dei punti chiave del Convegno, il programma coinvolge in un dialogo attivo la comunità educante, il pubblico, le istituzioni e le politiche della ricerca. Il progetto è quello di costruzione di una società democratica, in cui la comunicazione pubblica della scienza, l'alfabetizzazione digitale, contribuiscono a formare una cittadinanza scientifica digitale: accesso all'informazione e alla sua produzione, inclusione, trasparenza...

Il XV Convegno annuale AIUCD 2026 si propone di esplorare proprio il rapporto tra Digital Humanities e public engagement, offrendo spazio a modelli teorici, metodologie, servizi, applicazioni pratiche che favoriscono il coinvolgimento attivo dei cittadini nei processi di ricerca, documentazione nel campo delle Humanities. Sono temi di interesse del convegno le esperienze che, attraverso il digitale, attivano percorsi di co-creazione, crowdsourcing, narrazione digitale, musealizzazione, archiviazione virtuale e di citizen science, heritage sharing, dentro cornici Open Access e con approcci innovativi anche alla comunicazione scientifica. Metodi computazionali, sfide metodologiche e soluzioni sperimentali volte a migliorare accessibilità, interoperabilità, riuso e interpretabilità dei dati all'interno degli ecosistemi di ricerca digitale, per favorire la contaminazione e l'integrazione tra discipline e ambiti disciplinari e le pratiche di scienza aperta. AIUCD 2026 rappresenta un approfondimento di molteplici aspetti dell'ampio panorama delle Digital Humanities, con il fine di alimentare il confronto interdisciplinare e il dialogo scientifico in forma inclusiva.

La comunità degli umanisti digitali è stata invitata a contribuire al Convegno presentando proposte orali o in forma di poster intorno a 5 aree tematiche.

La prima area tematica, "DH e co-costruzione del sapere con le comunità: sfide, metodi, strumenti", è stata pensata per esplorare la partecipazione attiva di pubblici eterogenei mediante metodi collaborativi (crowdsourcing, citizen humanities, digital public history, pratiche di co-creazione, community museums ecc.) e strumenti digitali, inclusi sistemi basati sull'utilizzo dell'intelligenza artificiale, quest'ultimo tema importantissimo con il quale la comunità DH deve confrontarsi direttamente e riflettere. Questa area tematica include anche contributi che indagano i principi di citizen science in attività condivise di

produzione di conoscenza, come trascrizioni, annotazioni o mappature partecipate, così come lavori che favoriscono l'Open Access e l'adozione dei principi FAIR per garantire accessibilità, interoperabilità e riusabilità durevole di dati e risorse. Trovano particolare attenzione riflessioni e lavori che promuovono equità, accessibilità, cittadinanza digitale e inclusione di comunità marginalizzate tramite iniziative formative e pratiche educative partecipative in scuole, università, musei, MOOCs o laboratori territoriali.

La seconda area tematica, "Archivi ed edizioni: descrizioni aumentate, accessibilità e sistemi informativi", apparentemente più tradizionale, ha inteso focalizzarsi su riflessioni teoriche e applicative in merito alle opportunità e alle conseguenze dei processi di dematerializzazione della mediazione, al concetto di descrizione aumentata, intesa come opportunità di creare strumenti di mediazione capaci di integrare diverse entità in direzione dell'evoluzione dei sistemi informativi archivistici in sistemi interculturali, secondo le logiche che emergono da RiC e dagli standard di seconda generazione in genere. In questo senso trova particolare spazio il tema della rappresentazione e organizzazione a lungo termine del patrimonio culturale pregresso e dei materiali nativi digitali. Inevitabile in questo contesto una riflessione sulle strategie descrittive e sugli strumenti di restituzione delle informazioni, con riferimento alla creazione di risorse partecipative, alla realizzazione e gestione di edizioni critiche digitali e allo sviluppo di nuovi modelli di rappresentazione e visualizzazione dei dati che tengano conto delle opportunità di interazione e codifica orientata che la digitalità garantisce. Sono presenti proposte in tema di conservazione, gestione, sostenibilità e trasmissione di collezioni digitalizzate e nativamente digitali; progettazione di infrastrutture digitali resilienti e l'adozione di modelli sostenibili per la cura dei dati. Sono inoltre incluse presentazioni sulle buone pratiche e politiche per garantire l'accesso aperto, FAIR e a lungo termine, l'autenticità e la riusabilità delle risorse del patrimonio culturale digitale.

La terza area tematica, "Testualità digitali: prospettive, sviluppi e sperimentazioni", accoglie proposte che esplorano principalmente metodi e strumenti innovativi per lo studio computazionale dei testi, dai corpora letterari ai materiali del patrimonio culturale. Gli ambiti di interesse affrontati includono modelli computazionali per la critica letteraria, progressi nella linguistica computazionale e applicazioni di elaborazione del linguaggio naturale (NLP). Particolare attenzione è data a riflessioni ed esperienze che impiegano l'Intelligenza Artificiale e il Machine Learning per interpretare, modellare e visualizzare il patrimonio culturale, sia materiale che immateriale (ciò include anche fonti storiche, testi letterari). Analisi di rete, visualizzazione dei dati e tecniche di mappatura, soprattutto quando contribuiscono a superare i confini metodologici esistenti o aprono nuove prospettive interpretative sono temi di grande rilievo per la discussione.

La quarta area tematica, "Rappresentazione di dati e della conoscenza", intercetta ricerche dedicate alla rappresentazione della conoscenza e dell'uso delle tecnologie del Web Semantico all'interno delle discipline umanistiche digitali: progettazione di ontologie, modelli concettuali e grafi di conoscenza per i dati del patrimonio culturale, creazione di Linked Open Data e dataset conformi ai principi FAIR, e tecniche automatizzate di estrazione dei metadati, riconoscimento di entità nominate ed arricchimento semantico delle risorse culturali.

La quinta area tematica, "Memorie, storia e patrimoni culturali digitali", ha inteso approfondire principalmente la storia digitale, la co-generazione della memoria e il racconto del patrimonio culturale in digitale, incluse specifiche riflessioni su preservazione degli esiti delle ricerche e dei progetti scientifici digitali, anche nella prospettiva della scienza aperta (Open Science) e con applicazione dei principi FAIR. Inevitabile una riflessione sulle criticità infrastrutturali e sui problemi di sostenibilità dei progetti legati alla digitalizzazione e alle collezioni nate digitalmente, nonché proposte per la messa in atto di buone pratiche che garantiscano preservazione e accesso effettivo al patrimonio con abbattimento di impedimenti fisici, geografici, economici.

Al XV Convegno annuale AIUCD 2026 sono pervenute 151 proposte, 134 sono stati i contributi accettati: una comunità vivace e collaborativa, infatti si è avuta una media di 2,46 autori per contributo, 18 sono gli autori e i co-autori internazionali, a testimonianza delle collaborazioni internazionali e interdisciplinari della nostra comunità. I dati ci dicono, inoltre, quanto le Digital Humanities siano ormai presenti attivamente nel territorio nazionale, 163 tra Università, istituti CNR, centri e istituzioni di ricerca e culturali sono rappresentate al Convegno; Tutte le proposte sono state valutate attraverso un processo di revisione a doppio cieco, gestito da un Comitato Scientifico composto da 10 track chair e un Comitato di Programma di 94 revisori, oltre ai General Chairs.

I contributi, 95 relazioni orali e 39 poster, sono stati organizzati in panel per favorire lo scambio e la discussione su tematiche specifiche, due sessioni sono specificatamente dedicate alla presentazione dei poster.

Il Convegno è chiuso dall'incontro tra Philip Crowe, Direttore di Ricerca dell'UCD, School of Architecture, Planning and Environmental Policy, dell'University College di Dublino, che ha presentato una riflessione sul sostegno che le tecnologie digitali possono dare ad approcci basati sul territorio per affrontare le sfide sociali nelle città e contribuire a coinvolgere le comunità nel processo di cambiamento. Questa riflessione è oggetto di un dialogo tra Crowe e Arianna Ciula (King's College London) e Francesca Tomasi (Alma Mater Studiorum Università di Bologna)

Anche quest'anno il Convegno è stato preceduto da Aspettando AIUCD2026 che ha visto una stretta collaborazione tra l'Università di Cagliari con il DH UNICA - Centro interdipartimentale per l'Umanistica Digitale, i gruppi AIUCD Scuola (e il Liceo Statale Mamiani di Roma), Osservatorio DH AI, ALDiNA, e il Clarin Knowledge Center DiPText, e con il patrocinio di Wikimedia Italia e AIPH. La rassegna è stata curata, seguita e animata da Raffaele Argiolas, DH UniCa, Università degli Studi di Cagliari; Emmanuela Carbé, VeDPH, Università Ca' Foscari di Venezia; Federico Boschetti, CNR-ILC, Pisa - VeDPH, che ringraziamo per questa importante occasione di scambio

Dal punto di vista organizzativo questo XV Convegno annuale AIUCD 2026 è stato gestito da un Comitato Organizzativo multidisciplinare, che ha cercato di trovare soluzioni funzionali e innovative alla gestione del Convegno anche per favorire la socializzazione e lo scambio tra i partecipanti e, soprattutto, offrire degli strumenti che rendessero la proposta scientifica e il soggiorno a Cagliari fruibile, accessibile e condivisibile.

Per favorire il più possibile la fruizione del fitto programma, grazie ad Alessandro Capra, abbiamo potuto offrire ai nostri convegnisti un Companion del convegno, che raccoglie, in modo leggero e navigabile, tutto quello che serve sapere — dal programma minuto per minuto alle planimetrie delle aule, dalle 134 relazioni accettate ai luoghi più belli di Cagliari raggiungibili a piedi dalla sede. Companion opera insieme a Noa, un piccolo applicativo che propone ai convegnisti dei collegamenti tra panel, autori e sessioni, costruendo la propria personale scaletta di appuntamenti da seguire.

Concludiamo questa prefazione con un sentito ringraziamento a tutte le persone e gli enti che hanno reso possibile la realizzazione di AIUCD 2026. Un ringraziamento speciale va al Direttivo AIUCD, e in particolare alla presidente Marina Buzzoni, al segretario Paolo Monella e alla tesoriere Francesca Frontini, per il supporto istituzionale e la preziosa collaborazione in tutte le fasi preparatorie del convegno e nelle giornate Cagliaritanee. Siamo grati al Comitato Scientifico e al Comitato di Programma, per la disponibilità e l'attenzione nella valutazione delle proposte, così come a tutte le persone che hanno risposto alla call for papers e partecipato al convegno.

Un grazie sentito va al Comitato Organizzatore, per l'instancabile impegno, e la generosa disponibilità nella gestione di ogni fase del convegno.

Ringraziamo inoltre la direzione e il personale della segreteria amministrativa del Dipartimento di Lettere, Lingue e Beni culturali e della Facoltà di Studi Umanistici che hanno ospitato, sostenuto e co-finanziato il convegno e con loro gli studenti dell'Università di Cagliari che hanno offerto un prezioso contributo durante le tre giornate dell'appuntamento.

Infine, rinnoviamo la nostra gratitudine ai nostri sponsor: Presidenza della Regione Sardegna, Fondazione di Sardegna, Comune di Cagliari, all'ISSASCO e all'Istituto Parri, Associazione Iztokor, Clarin, che, in varie forme, hanno sostenuto AIUCD 2026 e ne hanno reso possibile la realizzazione.

In chiusura, ci piace pensare che la cittadinanza digitale non risiede nelle tecnologie che scegliamo di usare o meno, ma nelle comunità che nel digitale individuano un mezzo e una filosofia per dare forza a progetti culturali e scientifici pensati e realizzati per e con le persone.

Cagliari, maggio 2025

CRISTINA MARRAS, ANDREA PERGOLA, GIAMPAOLO SALICE

Legenda

Le 5 aree tematiche del Convegno sono indicate attraverso la seguente legenda



Archivi ed edizioni: descrizioni aumentate, accessibilità e sistemi informativi



DH e co-costruzione del sapere con le comunità: sfide, metodi, strumenti



Memorie, storia e patrimoni culturali digitali



Rappresentazione di dati e della conoscenza



Testualità digitali: prospettive, sviluppi e sperimentazioni



Altri contributi su Digital Humanities



Addestramento di un modello di HTR in Transkribus per testi plurilingui di età moderna: il caso del *Ripulimento della lingua sarda (Madau)*

Michela Incollu¹, Giulia Murgia²

¹Università degli Studi di Siena, Italy - m.incollu@student.unisi.it

²Università degli Studi di Cagliari, Italy - giulia.murgia@unica.it

ABSTRACT (ITALIANO)

Il contributo illustra le metodologie adottate per l'addestramento, tramite la piattaforma Transkribus, di un modello sperimentale di Handwritten Text Recognition (HTR) applicato a un manoscritto plurilingue di età moderna, il *Ripulimento della lingua sarda* di Matteo Madau (XVIII secolo). Viene riportato l'intero flusso di lavoro: preparazione dati e analisi del layout del documento; creazione della Ground Truth, con particolare attenzione alla metodologia impiegata per garantire la coerenza dei dati utili all'addestramento; criteri adottati per la selezione del modello base; avvio del processo di fine-tuning. I risultati del primo ciclo di addestramento sono presentati e valutati attraverso successivi test del modello su nuove pagine non viste durante l'addestramento. Le trascrizioni ottenute sono analizzate in relazione alla tipologia e alla frequenza degli errori di riconoscimento automatico. Tale analisi consente di misurare le effettive prestazioni del modello e di individuarne le principali criticità, in vista di futuri cicli di addestramento volti a migliorarne l'affidabilità.

Parole chiave: Transkribus; HTR; Ground Truth; fine-tuning; testo plurilingue.

ABSTRACT (ENGLISH)

Training HTR Text Model on Transkribus for Early Modern Multilingual Texts: the Case of Ripulimento della lingua sarda (Madau)

The present contribution illustrates the methodologies adopted for the training, on the Transkribus platform, of an experimental Handwritten Text Recognition (HTR) model, applied to a multilingual early modern handwritten text, *Ripulimento della lingua sarda* by Matteo Madau (18th century). The entire workflow is presented, including data preparation and document layout analysis, the creation of the Ground Truth, with particular attention to the methodology employed to ensure the coherence of the data used for training, the principles adopted for the selection of the base model and the initiation of the fine-tuning process. The results of the first training cycle are presented and evaluated through subsequent tests of the model on new pages not seen during training. The transcriptions obtained are analyzed in relation to the type and frequency of automatic recognition errors. This analysis allows for the evaluation of the actual performance of the model and to identify the major critical issues, useful for future training cycles aimed at improving its reliability.

Keywords: Transkribus; HTR; Ground Truth; fine-tuning; multilingual text.

1. INTRODUZIONE¹

Il presente contributo si propone di esplorare le metodologie coinvolte nell'addestramento, condotto sulla piattaforma Transkribus, di un modello sperimentale di HTR (Handwritten Text Recognition) basato sulla tecnica del fine-tuning incrementale a partire da un modello base pre-addestrato. L'obiettivo principale è riflettere sull'applicazione di tali metodologie a un testo manoscritto plurilingue, il *Ripulimento della lingua sarda* di Matteo Madau (XVIII secolo), e analizzare le differenze di apprendimento del modello nelle due sezioni principali del manoscritto (saggio e dizionario), caratterizzate da diversi gradi di complessità, sia strutturale che linguistica.

¹L'articolo è il frutto di una stretta collaborazione tra le autrici. Sono tuttavia da attribuire a Michela Incollu i parr.2,4,5,6 e a Giulia Murgia il par. 3. Il par. 1 è da attribuire ad entrambe le autrici.

Il progetto si inserisce in una prospettiva più ampia di valorizzazione del patrimonio linguistico e culturale sardo; trattandosi di un documento finora inedito, la sua digitalizzazione favorirebbe una fruizione più accessibile, anche da parte del pubblico non specializzato.

2. TRANSKRIBUS

Transkribus è una piattaforma di digitalizzazione e riconoscimento automatico del testo dotata di intelligenza artificiale, pensata per permettere anche a studiosi non esperti di coding e deep learning di accedere a strumenti avanzati di trascrizione automatica. Sviluppata nell'ambito del progetto READ dall'Università di Innsbruck in occasione del programma di finanziamento e sviluppo europeo Horizon 2020 (*Recognition and Enrichment of Archival Documents | READ | Project | Fact Sheet | H2020*, s.d.), a partire da un iniziale progetto denominato tranScriptorium (Sánchez et al., 2013), è attualmente gestita dalla cooperativa READ-COOP. La piattaforma si serve del motore di riconoscimento PyLaia, un toolkit che integra reti neurali convoluzionali (CNN) per l'estrazione delle caratteristiche visive, ossia i tratti grafici, e reti neurali ricorrenti (RNN), per la modellazione sequenziale del contesto testuale (Puigcerver, 2017/2026). Quest'architettura consente un apprendimento più stabile e veloce rispetto a sistemi HTR precedenti, basati su reti neurali ricorrenti di tipo Multidimensional Long Short-Term Memory (MDLSTM), caratterizzati da una maggiore complessità computazionale (Puigcerver, 2017). Un ulteriore vantaggio è costituito dal fatto che la piattaforma supporta numerose lingue, antiche e moderne, e, permettendo la riutilizzabilità dei modelli pubblici in processi di fine-tuning incrementale, consente l'addestramento di nuovi modelli personalizzati che si adattino a documenti redatti in lingue minoritarie non ancora rappresentate.

Una criticità della piattaforma Transkribus, se valutata sul confronto con soluzioni e ambienti operativi analoghi, riguarda la dipendenza da un sistema di crediti a pagamento, necessari per l'analisi e il riconoscimento automatico di testo e layout; questo sistema rende la piattaforma meno adatta a gestire progetti su larga scala, in quanto la dipendenza economica tende ad aumentare progressivamente. Esistono alternative gratuite e open-source, come l'infrastruttura eScriptorium (Kießling et al., 2019; Stokes et al., 2021), che permette, inoltre, l'esportazione e il riuso dei modelli HTR anche al di fuori dell'ambiente in cui sono stati prodotti. Transkribus, al contrario, permette il riuso di modelli pubblici solo all'interno del proprio ecosistema.

Tuttavia, l'integrazione "in-the-cloud" del motore PyLaia, offerta direttamente all'interno della piattaforma, rappresenta un notevole vantaggio di utilizzo; a differenza di eScriptorium, infatti, non è richiesta alcuna installazione locale, che richiederebbe all'utente competenze tecniche per la gestione dell'infrastruttura, né la disponibilità di risorse computazionali particolarmente avanzate.

In linea con le pratiche relative al public engagement, la piattaforma Transkribus riunisce una vasta comunità di studiosi e incoraggia la condivisione dei modelli HTR addestrati, in particolare per scritti e lingue sotto-rappresentate. In quest'ottica, l'addestramento di un modello HTR su un documento come il *Ripulimento della lingua sarda* e la sua successiva pubblicazione all'interno della sezione dedicata ai modelli condivisi costituirebbero un contributo importante, agevolando contestualmente future trascrizioni di manoscritti plurilingui.

3. DESCRIZIONE MANOSCRITTO

Il *Ripulimento della lingua sarda* di Matteo Madau è l'opera che, nel XVIII secolo, dà avvio agli studi di linguistica sarda (Murgia, 2023). L'autore diede alle stampe un primo *Saggio* dell'opera nel 1782 (Cagliari, Titard). Tuttavia, del *Ripulimento* esiste anche un manoscritto ad oggi inedito, scritto di pugno dallo stesso Madau, suddiviso in due tomi cartacei (rispettivamente di 196 e 293 carte), oggi appartenente al Fondo Baille della Biblioteca Universitaria di Cagliari, dove è conservato con segnatura S.P.6.1.39 e S.P.6.1.40. Rispetto al *Saggio* a stampa, il *Ripulimento* inedito manoscritto è l'edizione integrale che contiene in maniera compiuta la riflessione linguistica di Madau: i due tomi che lo compongono accolgono non solo il discorso introduttivo sulla lingua sarda, ma anche un ricco dizionario storico-etimologico bipartito in due sezioni, con ventimila voci sarde di presunta derivazione greca e ben centomila vocaboli derivati dal latino.

Le carte hanno una dimensione di 355 x 223 mm e sono fittamente scritte a piena pagina. Il numero di righe è mutevole e oscilla tra le 23 e le 51. Il testo appare redatto da un'unica mano.

Nel primo tomo, la sezione che precede il dizionario – contenente la dedica al re Vittorio Amedeo III, le indicazioni dell'autore a chi legge, l'introduzione, la presentazione dell'opera e le indicazioni ortografiche –

presenta un testo disteso su un'unica colonna, a piena pagina (cc. 1r-42r). Come si può notare dalla figura 1, lo spazio riservato ai margini destro e sinistro accoglie le note apposte da Madau, richiamate a testo mediante un esponente alfabetico. A rendere affollate le pagine contribuisce anche l'assenza di spazi tra un paragrafo e l'altro. Non mancano poi interventi dovuti a ripensamenti dell'autore, come cancellature e aggiunte in interlinea. Abbondante inoltre è l'uso della sottolineatura, impiegata per segnalare le citazioni in lingue diverse dall'italiano e le parole che, in una futura edizione a stampa mai realizzata, sarebbero state da rendere in corsivo.



Figura 1. Ripulimento della lingua sarda.
Cagliari, Biblioteca Universitaria, Fondo Baille, S.P.6.1.39, cc. 10v-11r
Su concessione del Ministero della Cultura / Biblioteca Universitaria di Cagliari
 A partire dalla c. 43r del primo tomo, inizia la sezione del *Dizionario di sardi voci, portate dal Greco*, nella quale i lemmi ordinati alfabeticamente si distendono su due colonne.



Figura 2. Ripulimento della lingua sarda.
Cagliari, Biblioteca Universitaria, Fondo Baille, S.P.6.1.39, cc. 45v-46r
Su concessione del Ministero della Cultura / Biblioteca Universitaria di Cagliari
 Le figg. 1 e 2 rendono chiare le principali difficoltà che si pongono al riconoscimento automatico del nostro testo: la presenza di una lingua come il sardo, che non è contemplata tra i modelli proposti dalla piattaforma perché mai messa alla prova su Transkribus; la presenza di layout eterogenei (organizzazione dello specchio di scrittura su una o due colonne; presenza di glosse nei margini; uso della sottolineatura; ecc.); l'alternarsi di più alfabeti (latino e greco); il plurilinguismo dell'opera, scritta in italiano, ma con una cospicua presenza di sardo, latino e greco.

4. PREPARAZIONE DEI DATI²

La preparazione dei dati costituisce una fase fondamentale nel processo di addestramento di un nuovo modello di riconoscimento automatico del testo.

Per prima cosa vengono caricati, all'interno di una nuova collezione, i documenti da analizzare, in formato jpg e alla risoluzione di 300 dpi; la digitalizzazione del manoscritto di Madau è stata fornita dalla Biblioteca Universitaria di Cagliari, che, attraverso il servizio DIGIT, fornisce su richiesta le copie digitali per uso di studio o di ricerca.

Una volta completato il caricamento, occorre eseguire l'analisi e il riconoscimento del layout; questa fase preliminare, infatti, consente al software di individuare le aree testuali (sezioni e righe di testo) e di analizzarne i tratti grafici, così da apprendere le corrispondenze tra immagine e trascrizione.

Sulla piattaforma è possibile automatizzare l'analisi del layout, selezionando le pagine interessate e scegliendo, tra i modelli disponibili, quello che più si adatta alle caratteristiche del documento in esame.

Nel nostro caso sono stati inizialmente presi in considerazione due modelli: *Mixed Line Orientation* e *Universal Lines*. La scelta è stata condizionata dalla struttura eterogenea del documento, caratterizzato dalla compresenza di una prima sezione con layout abbastanza regolare (saggio) e di una seconda, più cospicua, organizzata in testo a due colonne (dizionario), oltre alla presenza di numerose note a margine, esterne alla cornice di testo principale.

Sulla base di tali caratteristiche, il modello *Mixed Line* è apparso inizialmente l'opzione migliore, in quanto addestrato a gestire layout eterogenei; tuttavia, si è preferito optare per il modello *Universal Lines*, da applicare all'intero corpus in modo da ridurre il rischio di over-segmentation e ottenere una base più stabile su cui condurre le analisi successive.

Entrambi i modelli hanno mostrato da subito significative difficoltà a individuare correttamente il layout nella sezione con testo a due colonne, compromettendo così la segmentazione e alterando l'ordine naturale delle righe di testo. Per ovviare a queste difficoltà è stato avviato l'addestramento di un modello *Baseline* specifico per il layout del documento. Il nuovo modello (*Baseline Model Madau - experimental*), addestrato su un campione di pagine preparate tramite la selezione manuale delle aree testuali, ha prodotto risultati soddisfacenti solo se applicato a singole pagine, mentre si mostra ancora inefficace nella visione a doppia pagina che caratterizza le riproduzioni digitali del manoscritto. Necessita, pertanto, di ulteriori cicli di addestramento, in modo da poter raggiungere la soglia di *loss* inferiore al 10%; allo stato attuale, infatti, registra un valore pari a 14,84%.

Un'ulteriore fase centrale è rappresentata dalla preparazione della Ground Truth, ossia la raccolta di dati, verificati accuratamente, utilizzata per condurre l'addestramento, la convalida e infine il test dei modelli di riconoscimento automatico del testo. Secondo le linee guida Transkribus, reperibili all'interno della sezione *Training Text Models*, alla voce *Data Preparation*, per una Ground Truth consistente è necessario trascrivere manualmente un numero di pagine che varia a seconda delle caratteristiche del testo in esame. Nel caso di testi manoscritti, in particolare, si consiglia di raccogliere almeno 10.000 parole per ogni mano così da condurre un addestramento che possa raggiungere potenzialmente un *Character Error Rate* (CER) compreso tra il 2% e il 4%.

Un altro aspetto importante nella costruzione della Ground Truth riguarda l'accuratezza della trascrizione e la coerenza delle scelte editoriali adottate; si tratta di elementi indispensabili per un apprendimento più veloce ed efficiente del modello, in quanto, come evidenziato da studi recenti, la qualità della Ground Truth ha un impatto significativo sull'addestramento, e di conseguenza sulle prestazioni, dei modelli HTR. Significativamente, è stato dimostrato che una Ground Truth di dimensioni ridotte ma accurata e coerente contribuisca a minimizzare l'insorgenza di errori sistematici, che tendono invece a manifestarsi con maggiore frequenza laddove i modelli siano stati addestrati su una gran mole di dati approssimativi (Jungo et al., 2023).

² La metodologia impiegata per l'avvio e lo sviluppo del lavoro su Transkribus si fonda principalmente sulle linee guida ufficiali disponibili all'interno della piattaforma stessa, in particolare alla sezione contenente i materiali relativi all'addestramento dei modelli di riconoscimento automatico del testo (Training Text Recognition Models, s.d.), e alle pagine dedicate alla preparazione dei dati (Data Preparation, s.d.) e alla descrizione dei modelli (Model Setup and Training, s.d.).

Per ragioni di tempo si è scelto di concentrare il lavoro di trascrizione su un sottoinsieme rappresentativo di 20 pagine (22.576 parole, 2.444 righe),³ tratte dal primo tomo del manoscritto di Madau. Le pagine sono state scelte in maniera casuale, in modo da raccogliere una quantità quanto più ampia di dati testuali, e ripartite in egual numero tra le due sezioni principali (saggio e dizionario). Le trascrizioni sono state quindi salvate e marcate come Ground Truth.

5. ADDESTRAMENTO

Completata la preparazione dei dati di Ground Truth, è stato possibile procedere con l'addestramento di un nuovo modello di riconoscimento automatico. Procedendo all'interno della collezione in esame, vengono selezionate le pagine precedentemente trascritte e marcate come Ground Truth, scegliendo dal menù a tendina la voce *New Train Model*. Le pagine vengono quindi suddivise automaticamente in due gruppi: training set, su cui avverrà l'addestramento vero e proprio, e validation set, corrispondente, secondo lo standard espresso dalle direttive di Transkribus, al 10% dei dati totali, utilizzato per monitorare le prestazioni del modello durante l'addestramento.

Successivamente vengono richiesti i metadati descrittivi relativi al nuovo modello: nome (Madau multilingual model [experimental]), breve descrizione (Experimental model for 18th century plurilingual handwritten text trained on Matteo Madau's work, "Ripulimento della lingua sarda"), periodo cronologico (XVIII secolo) e lingue presenti nel documento (Italian; Sardinian; Latin; Greek, ancient). All'interno della sezione riguardante i parametri avanzati, inoltre, ogni utente può verificare, ed eventualmente modificare, le impostazioni automatiche riguardo *Epochs* (numero di volte in cui l'algoritmo di apprendimento esaminerà i dati di addestramento e si valuterà sia sul training set che sul validation set), e *Early Stopping* (numero minimo di epochs per addestramento).

Una volta compilato il pannello relativo ai metadati, viene scelto il modello di base che verrà usato dal motore di riconoscimento di testo come punto di partenza per addestrare il nuovo modello, secondo il processo di *fine-tuning*. È stato selezionato un modello che potesse avvicinarsi maggiormente alle caratteristiche del documento, filtrando i diversi modelli disponibili per tipologia (handwriting) e per lingua e privilegiando modelli addestrati su testi multilingui. La scelta è ricaduta sul modello denominato *Italian 16th century antiquarian hand - Pirro Ligorio 0.3*, addestrato su un corpus plurilingue (latino, greco, italiano) di 57.003 parole tratte dall'*Enciclopedia dell'Antichità*, opera di carattere enciclopedico (Il modello HTR, s.d.; *Italian 16th century antiquarian hand - Pirro Ligorio 0.3*, s.d.). Queste caratteristiche, unite al tasso di errore CER del 2,70%, rendono il modello adatto a costituire una base valida per l'addestramento del nuovo modello. Inoltre, come evidenziato da recenti studi sperimentali condotti su corpora plurilingui, un singolo modello HTR multilingue costituisce un'alternativa efficace e conveniente a singoli modelli monolingui, soprattutto nel caso in cui le diverse sezioni linguistiche non siano facilmente individuabili e separabili in maniera automatica (Capurro et al., 2023). Questo permette di ovviare in maniera efficiente alla mancanza di un modello di riconoscimento specifico per la lingua sarda, che, in quanto lingua minoritaria, non è attualmente rappresentata nel panorama linguistico di Transkribus. A questo punto è stato possibile proseguire con l'avvio del training vero e proprio.

Durante il primo ciclo di addestramento, l'attività di apprendimento viene svolta sul training set e, progressivamente, le prestazioni del modello vengono valutate mediante il calcolo del Character Error Rate (CER), effettuato sui dati di convalida del validation set. Il processo di addestramento si conclude automaticamente quando il modello non mostra più ulteriori miglioramenti, ossia quando viene raggiunto il valore minimo di CER.

Nel nostro caso, al termine del primo ciclo di training, il modello ha raggiunto un CER del 3,84%. Si tratta di un risultato incoraggiante, soprattutto in questa fase iniziale, in quanto da una parte conferma la solidità del modello di base scelto per il fine-tuning, e dall'altra lascia ben sperare in un buon potenziale di miglioramento, da verificarsi nelle successive fasi di addestramento incrementale. La figura 3 riporta il grafico della curva di apprendimento (Learning Curve), generata alla fine del primo ciclo di addestramento, che mostra la progressiva diminuzione del CER all'aumentare del numero di epoch.

³ Poiché ciascuna riproduzione fotografica mostra il volume aperto su due facciate, tale selezione corrisponde complessivamente a 40 pagine manoscritte.

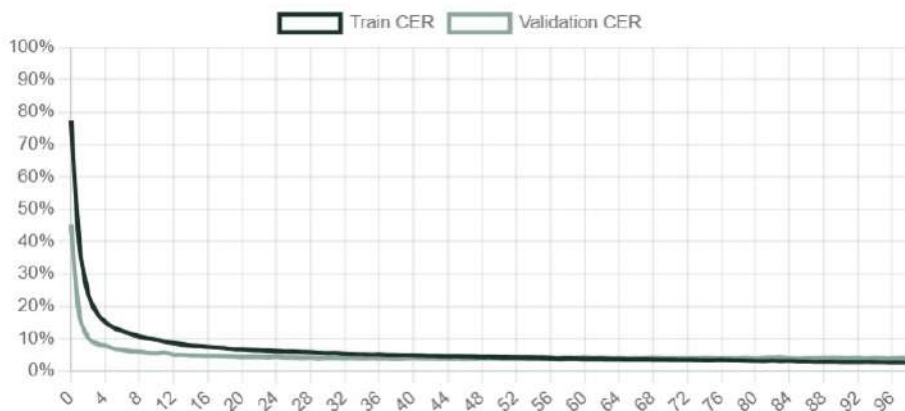


Figura 3. Grafico della *Learning curve* alla fine del primo ciclo di addestramento.

La fase successiva prevede di applicare il modello addestrato anche ad altre pagine mai viste. Questo permette sia di testare la reale capacità di generalizzazione del modello su dati non elaborati in precedenza, sia di ampliare progressivamente la Ground Truth, in modo da poter proseguire l’addestramento su un dataset sempre più ampio, con l’obiettivo di aumentare il tasso di attendibilità e ridurre ulteriormente il CER.

6. RISULTATI

Applicando il riconoscimento automatico a un insieme di pagine non incluse nell’iniziale Ground Truth, sono state generate nuove trascrizioni, successivamente analizzate mediante confronto diretto con le immagini del manoscritto. Da una prima analisi quantitativa degli errori è emersa una media di tre errori per riga, corrispondente a un tasso di errore per carattere prossimo al 3%, coerente con il CER ottenuto in fase di addestramento. Dal punto di vista qualitativo gli errori individuati si dividono principalmente in errori dovuti alla segmentazione non ottimale del layout o alla qualità dell’immagine, compromessa in casi sporadici da macchie e rumore generico, ed errori ortografici, riguardanti la resa delle maiuscole o il riconoscimento delle vocali accentate.

Le sezioni che presentano le maggiori criticità si confermano quelle caratterizzate da frequenti cambi di lingua, in cui si alternano, all’interno della stessa riga, parole in caratteri latini, riconducibili a sardo, latino o italiano, e termini in greco. Questa tendenza appare indicativa dei limiti intrinseci di un’architettura, come quella di PyLaia, che, pur dimostrandosi solida e veloce nell’apprendimento visivo-sequenziale, mostra una minore efficacia nella modellazione di contesti linguistici variabili, confrontata con i risultati dei più recenti Transformer, come TrOCR, secondo quanto emerge da recenti studi comparativi sulle moderne tecnologie HTR (Romein et al., 2025).

Nonostante ciò, il modello evidenzia un buon livello di maturità già dopo il primo ciclo di addestramento, pur richiedendo ancora una revisione sistematica delle trascrizioni.

Ulteriori cicli di addestramento incrementale potranno migliorare le prestazioni e garantire trascrizioni sempre più affidabili ed esportabili. La criticità principale resta legata all’analisi del layout, che richiede un’attenzione particolare e un addestramento mirato di un modello specifico per la sezione del dizionario, caratterizzato da una struttura più eterogenea e complessa.

Trattandosi di un’opera di notevole rilievo storico-linguistico e considerata la cospicua estensione del testo, l’avviamento di un progetto di digitalizzazione di questo tipo faciliterebbe la realizzazione di un’edizione critica integrale del manoscritto. L’esportazione delle trascrizioni in formato TEI XML permette infatti la condivisione dei dati nonché il loro uso in differenti contesti di ricerca e disseminazione, favorendo una fruizione più accessibile e partecipata anche al di fuori della comunità accademica, in una prospettiva di public engagement capace di rispondere alle odierne esigenze della ricerca nell’ambito dell’umanistica digitale (Pierazzo, 2016).

BIBLIOGRAFIA

- Capurro, C., Provatorova, V., & Kanoulas, E. (2023). Experimenting with Training a Neural Network in Transkribus to Recognise Text in a Multilingual and Multi-Authored Manuscript Collection. *Heritage*, 6(12), 7482–7494. <https://doi.org/10.3390/heritage6120392>
- Data Preparation. (s.d.). Recuperato 27 gennaio 2026, da <https://help.transkribus.org/data-preparation>
- How to train a model in Transkribus. (2025, ottobre 22). <https://blog.transkribus.org/en/how-to-train-a-text-recognition-model-in-transkribus>
- Il modello HTR. (s.d.). LIGORIO DIGITALE 1.0 WORKSPACE. Recuperato 2 febbraio 2026, da <https://lidiws-limes.cfs.unipi.it/trascrizione/ilmodellohtr/>
- Italian 16th century antiquarian hand—Pirro Ligorio 0.3. (s.d.). Recuperato 4 febbraio 2026, da <https://app.transkribus.org/models/public/text/42105>
- Jungo, M., Vögtlin, L., Fakhari, A., Wegmann, N., Ingold, R., Fischer, A., & Scius-Bertrand, A. (2023). Impact of Ground Truth Quality on Handwriting Recognition. *Proceedings of the 12th International Symposium on Information and Communication Technology*, 135–140. <https://doi.org/10.1145/3628797.3628976>
- Kiessling, B., Tissot, R., Stokes, P., & Stökl Ben Ezra, D. (2019). eScriptorium: An Open Source Platform for Historical Document Analysis. *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, 19–19. <https://doi.org/10.1109/ICDARW.2019.10032>
- Model Setup and Training. (s.d.). Recuperato 27 gennaio 2026, da <https://help.transkribus.org/model-setup-and-training>
- Murgia, G. (2023). «Tutto ciò è necessario alla Sarda nazione non men che a tutte le altre dell'Europa». La tradizione manoscritta e a stampa del Ripulimento della lingua sarda di Matteo Madau. Il «traffico delle lingue». *Idiomi a contatto in Sardegna e nel Mediterraneo in età preunitaria*, eds. R. Fresu, P. Maninchedda, G. Murgia, P. Serra, Cagliari, UNICApres, 2023, pp. 491-529. ISBN: 978-88-3312-112-3. <https://doi.org/10.13125/UNICAPRESS.978-88-3312-108-6>
- Pierazzo, E. (2016). 3. **Modelling** Digital Scholarly Editing: From Plato to Heraclitus. In M. J. Driscoll & E. Pierazzo (A c. Di), *Digital Humanities Series* (1ª ed., Vol. 4, pp. 41–58). Open Book Publishers. <https://doi.org/10.11647/OBP.0095.03>
- Puigcerver, J. (2017). Are Multidimensional Recurrent Layers Really Necessary for Handwritten Text Recognition? *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 67–72. <https://doi.org/10.1109/ICDAR.2017.20>
- Puigcerver, J. (2026). *Jpuigcerver/PyLaia* [Python]. <https://github.com/jpuigcerver/PyLaia> (Opera originale pubblicata 2017)
- PyLaia: Teklia Documentation. (s.d.). Recuperato 27 gennaio 2026, da <https://doc.tekليا.com/pylaia/>
- Recognition and Enrichment of Archival Documents | READ | Project | Fact Sheet | H2020. (s.d.). [CORDIS | European Commission. Recuperato 4 febbraio 2026, da https://cordis.europa.eu/project/id/674943](https://cordis.europa.eu/project/id/674943)
- Romein, C. A., Rabus, A., Leifert, G., & Ströbel, P. B. (2025). Assessing advanced handwritten text recognition engines for digitizing historical documents. *International Journal of Digital Humanities*, 7(1), 115–134. <https://doi.org/10.1007/s42803-025-00100-0>
- Sánchez, J. A., Mühlberger, G., Gatos, B., Schofield, P., Depuydt, K., Davis, R. M., Vidal, E., & de Does, J. (2013). tranScriptorium: A european project on handwritten text recognition. *Proceedings of the 2013 ACM symposium on Document engineering, DocEng '13*, 227–228. <https://doi.org/10.1145/2494266.2494294>
- Stokes, P. A., Kiessling, B., Stökl Ben Ezra, D., Tissot, R., Gargem, E. H. (2021). The eScriptorium VRE for Manuscript Cultures. *Classics@*, 2021, 18 (1), <https://classics-at.chs.harvard.edu/classics18-stokes-kiessling-stokl-ben-ezra-tissot-gargem/>. (hal-03991423)
- Tarride, S., Schneider, Y., Generali-Lince, M., Boillet, M., Abadie, B., & Kermorvant, C. (2024). Improving Automatic Text Recognition with Language Models in the PyLaia Open-Source Library (arXiv:2404.18722; Versione 1). *arXiv*. <https://doi.org/10.48550/arXiv.2404.18722>
- Training Text Recognition Models. (s.d.). Recuperato 27 gennaio 2026, da <https://help.transkribus.org/training-text-recognition-models>