

Atti del XIV Convegno Annuale

Diversità, Equità e Inclusione: Sfide e Opportunità per l'Informatica Umanistica nell'Era dell'Intelligenza Artificiale

Verona :: 11-13 giugno 2025

A cura di:

Simone Rebora • Marco Rospocher • Stefano Bazzaco



**UNIVERSITÀ
di VERONA**
Dipartimento
di LINGUE
E LETTERATURE STRANIERE



ASSOCIAZIONE per
l'INFORMATICA UMANISTICA
e la CULTURA DIGITALE

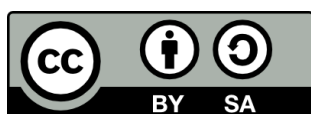


ISBN 978-88-942535-9-7



Copyright ©2025 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: Simone Reborà; Marco Rospocher; Stefano Bazzaco (2025). Diversity, Equity, and Inclusion: Challenges and Opportunities for Digital Humanities in the Age of Artificial Intelligence, Proceedings del XIV Convegno Annuale AIUCD, Verona 11-13 giugno 2025, Università di Verona.

Ultimo accesso agli URL in data 8 maggio 2025.

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 della conferenza AIUCD 2025 è disponibile online

<https://aiucd2025.dlss.univr.it/detailed-schedule/>

The AIUCD 2025 Conference Program is available online

<https://aiucd2025.dlss.univr.it/en-gb/detailed-schedule/>

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 2025.

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 2025.

Gli atti del convegno AIUCD 2025 sono pubblicati come raccolta di contributi in formato PDF forniti direttamente dagli autori e dalle autrici. I file sono stati raccolti e assemblati senza interventi redazionali da parte dei curatori.

The proceedings of the AIUCD 2025 conference are published as a collection of PDF contributions provided directly by the authors. The files have been collected and compiled without editorial intervention by the editors..

Il logo di AIUCD 2025 include l'immagine "Verona Dark Line Simple Minimalist Skyline With White Background" di @pabloprat/stock.adobe.com, ottenuta tramite la licenza Adobe Stock dell'Università di Verona.

The AIUCD 2025 logo includes the image "Verona Dark Line Simple Minimalist Skyline With White Background" by @pabloprat/stock.adobe.com, used under the Adobe Stock license of the University of Verona.

Il background della copertina è stato creato con tecniche di AI generativa con lo strumento "Magic Media" disponibile su Canva, usando un prompt con il tema del convegno.

The background of the cover was created using generative AI techniques with the "Magic Media" tool available on Canva, using a prompt based on the conference theme.

Comitato Organizzatore / *Organizing Committee*

General Chairs

Simone Rebora (Università degli Studi di Verona)
Marco Rospocher (Università degli Studi di Verona)

Local Chair

Anna Cappellotto (Università degli Studi di Verona)

Registration Chair

Giorgia Pomarolli (Università degli Studi di Verona)

Proceedings Chair

Stefano Bazzaco (Università degli Studi di Verona)

Sponsorship Chair

Matteo Lissandrini (Università degli Studi di Verona)

Publicity Chair

Sabrina Piccinin (Università degli Studi di Verona)

Comitato Scientifico / *Scientific Committee*

Program Chairs

Simone Rebora (Università degli Studi di Verona)
Marco Rospocher (Università degli Studi di Verona)

Digital Humanities e inclusione / *Inclusive DH*

Stefano Bazzaco (Università degli Studi di Verona)
Massimo Salgaro (Università degli Studi di Verona)

Archivi ed Edizioni Digitali / *Archives and Digital Editions*

Elisa Cugliana (Cologne Center for eHumanities)
Christian D'Agata (Università di Catania)

Metodi Computazionali / *Computational Methods*

Rachele Sprugnoli (Università degli Studi di Parma)
Sara Tonelli (Fondazione Bruno Kessler)

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

Francesco Mambrini (Università Cattolica del Sacro Cuore)
Elena Spadini (Universität Bern)

Preservazione della Memoria e del Patrimonio Digitale / *Preservation of Memory and Digital Cultural Heritage*

Monica Berti (Universität Leipzig)
Daria Spampinato (Istituto di Scienze e Tecnologie della Cognizione-CNR)

Comitato di programma / *Program committee*

Stefano Allegrezza (Università di Macerata); Laura Antonietti (Université de Versailles Saint Quentin en Yvelines (Université Paris-Saclay)); Luigi Bambaci (École pratique des hautes études, PSL); Liborio P. Barbarino (Università di Catania); Nicola Barbuti (Università degli Studi di Bari Aldo Moro (Dipartimento di Ricerca e Innovazione Umanistica)); Sofia Baroncini (Leibniz Institute of European History); Andrea Bellandi (Institute for Computational Linguistics (CNR)); Mario A. Bochicchio (University of Bari, Dep.t of Computer Science); Andrea Bolioli (Independent researcher); Marco Bombieri (University of Verona); Paolo Bonora (Università di Bologna); Flavia Bruni (Università di Chieti-Pescara); Marina Buzzoni (Ca' Foscari University of Venice); Alberto Campagnolo (KU Leuven); Anna Cappellotto (Università di Verona); Emanuela Carbé (Università di Siena); Vittore Casarosa (ISTI-CNR); Raffaele Cioffi (Università di Napoli Federico II); Fabio Ciotti (Università di Roma Tor Vergata); Vincenzo Colaprice (University of Turin); Giuseppe Consolo (Università degli studi di Napoli, Federico II); Elisa Conti (Università di Catania); Salvatore Cristofaro (CNR ISTC); Giulia D'Agostino (TU Darmstadt); Elisa D'Argenio (HUN-REN Hungarian Research Centre for Linguistics); Enrico Daga (The Open University); Stefano Dall'Aglio (Università Ca' Foscari Venezia); Marilena Daquino (University of Bologna); Mauro De Bari (University of Bari Aldo Moro); Angelo M. Del Grosso (CNR-ILC); Matteo Di Franco (Università di Napoli Federico II); Giorgio Maria Di Nunzio (University of Padua); Stefano Ferilli (University of Bari); Lorenzo Ferroni (Università degli Studi di Verona); Franz Fischer (Ca' Foscari Università Ca' Foscari); Greta H. Franzini (Eurac Research); Francesca Frontini (CNR-ILC); Daniele Fusi (VeDPH, Stuttgart University); Mariangela Giglio (University of Bologna); Tiago Luis Gil (University of Brasilia); Luca Giovannini (University of Potsdam); Milena Giuffrida (Università di Catania); Edmondo Grassi (Università Telematica San Raffaele Roma); Miryam Grasso (Università di Catania); Piergiovanna Grossi (Università di Verona); Fahad Khan (CNR-ILC); Michele Lacriola (Università di Siena); Maurizio Lana (Univ. del Piemonte Orientale); Federica Lazzerini (Università degli Studi di Torino); Eleonora Litta (Università Cattolica del Sacro Cuore, Milano); Dominique Longrée (ULiège); Diego Mantoan (University of Palermo); Anna Maria Marras (University of Turin); Cristina Marras (CNR); Pietro Mazzarisi (University of Trieste); Barbara McGillivray (King's College London); Federico V. Meschini (Tuscia University); Alessio Miaschi (Istituto di Linguistica Computazionale "A. Zampolli" (CNR-ILC), Pisa); Andrea Micheletti (University of Padua); Giulia Miglietta (Università del Salento); Paolo Monella (Università Kore di Enna); Johanna Monti (Università degli Studi di Napoli "L'Orientale"); Rossana Morriello (Università degli Studi di Firenze); Gloria Mugelli (ILC CNR); Serge Noiret (AIPH (Associazione Italiana di Public History) (European University Institute)); Giuseppe Palazzolo (Università di Catania); Mafalda Papini (CNR-ILC); Enrico Pasini (UniTO/CNR-ILIESI); Giulia Pedonese (CNR (Istituto di Linguistica Computazionale "Antonio Zampolli")); Paola Peratello (Università Ca' Foscari Venezia); Federico Pianzola (University of Groningen); Chantal Pivetta (Lund University (Sweden)); Igor Pizzirusso (AIPH); Giulia Re (ILC-CNR); Giulia Renda (University of Bologna); Dario Rodighiero (University of Groningen); Roberto Rosselli Del Turco (Università di Torino); Enrica Salvatori (Università di Pisa); Emilio M. Sanfilippo (CNR); Eva Sassolini (CNR-ILC); Andrea Schimmenti (University of Bologna); Flavia Sciolette (CNR-ILC); Pietro Sichera (ILIESI-CNR); Daniele Silvi (Università di Roma 'Tor Vergata'); Giulia Speranza (University of Naples L Orientale); Francesco V. Stella (UNISI); Timothy Tambassi (Ca' Foscari University of Venice); Mirko Tavosanis (Università di Pisa); Francesca Tomasi (Università di Bologna); Simona Turbanti (University of Milano); Marco Venuti (Università di Catania); Gennaro Vessio (University of Bari Aldo Moro); Gabriele Vezzani (Università di Verona / RWTH Aachen University); Fabio Vitali (University of Bologna).

Enti organizzatori / *Organisers*

AIUCD; Università di Verona: Dipartimento di Lingue e Letterature Straniere; Digital Arena for Inclusive Humanities (DAIH).

Sommario

Prefazione

Simone Reborà, Marco Rospocher, Stefano Bazzaco

I-II

Digital Humanities and Inclusion

PrevNet. A FAIR and inclusive resource for the study of proverbs in historical languages <i>Andrea Farina, Barbara McGillivray</i>	2
Potential bias in AI: cultural representation and the marginalization of African art <i>Francesca Bignotti</i>	11
Exploring data-driven narratives in Digital Humanities web-based projects: features and impact <i>Tommaso Battisti, Marilena Daquino</i>	18
Il Glossario delle Infrastrutture di Ricerca (GIR) <i>Lucia Francalanci, Alessia Scognamiglio, Irene Falini, Pietro Restaneo, Giulia Pedonese, Alessia Spadi</i>	24
Educational Impact of Storytelling and Data Visualisation in the Interpretation of Humanities Data <i>Giulia Renda, Marilena Daquino</i>	29
IncluInstIT: Un nuovo corpus per lo studio di linguaggio inclusivo su Instagram <i>Irene Caiazza, Giovanna Maria Dimitri, Liana Tronci</i>	35
Per un'analisi della rivista Umanistica Digitale in ottica DEI <i>Rossana Morriello, Lucia Sardo</i>	40
Carpe bias, quam minimum credula queries <i>Sabato Danzilli</i>	46
Evaluating bias within an epistemological framework for AI-based research in the humanities <i>Sarah Oberbichler, Cindarella Petz</i>	52
Semplificare la lettura dei manoscritti utilizzando tecnologie WEB interattive e interazioni «hover» <i>Giacomo Marchioro, Andrea Brugnoli, Francesca Carnazzi, Paolo Pellegrini, Edoardo Ferrarini</i>	60
Accessibilità e inclusione per la documentazione del restauro: gli archivi del Centro Conservazione e Restauro «La Venaria Reale» <i>Stefania De Blasi, Edi Guerzoni, Chiara Pipino</i>	66
Il patrimonio culturale digitale delle minoranze etniche: il progetto DIGICHer tra le comunità Sámi, ladine ed ebraiche in Europa <i>Matteo Cova, Eleonora De Longis</i>	72
Fulfilling GEN-der AImS: do image-generating tools discriminate? An on-field study <i>Francesco Meledandri</i>	78
From Bias Paralysis to Bias as a Category of Analysis. Introducing the Bias-Aware Framework <i>Mrinalini Luthra, Amber Zijlma</i>	86
Supporting Children with Linguistic Vulnerabilities Through Advanced, Theory-Driven Technological Solutions: The TELMI Approach for Italian children with DLD and Children with Italian as L2 <i>Arianna Compostella, Giulia Valcamonica, Mattia Gianotti, Matteo Secco, Silvia Silleresi, Fabrizio Arosio, Franca Garzotto, Maria Teresa Guasti</i>	94
Tecnologie AI per la didattica <i>Gabriele Prospero, Giulia Miglietta, Eleonora Miccoli, Mario Bochicchio</i>	104

DEA - An Innovative Technological Tool for Personalized Linguistic Training for Italian Children with Developmental Dyslexia <i>Marta Tagliani, Maria Vender, Giulia Valcamonica, Giovanni Caleffi, Franca Garzotto, Denis Delfitto</i>	110
Il ruolo delle Infrastrutture nella costruzione di un ambiente di ricerca inclusivo. Un modello di buone pratiche <i>Marta Caradonna, Nicola Giampietro, Roberta Bianca Luzietti, Monica Monachini, Valeria Quochi, Emiliano Degl'Innocenti, Alessia Spadi, Alessandra Caravale, Antonio D'Eredità, Paola Moscati, Giacomo Mancuso</i>	118
Predicting Grammatical Cases in Slovenian Varieties in Italy: A Use Case from the LORIS 1.1 Language Assistant <i>David Bordon</i>	124
Verso un futuro senza barriere: l'accessibilità dei documenti elettronici nell'European Accessibility Act <i>Stefano Allegrezza</i>	129
La gestione del nuovo sapere digitale contemporaneo. Scenari, criticità, sfide, prospettive <i>Nicola Barbuti</i>	134

Archives and Digital Editions

Learner Corpus of Creative Writing: An interdisciplinary challenge <i>Ioanna Tyrou, Katerina Florou</i>	143
Retrieval-Augmented Generation systems for enhanced access to digital archives <i>Michele Ciletti</i>	149
Preserving Clarity: The MAGIC project approach to ancient manuscripts <i>Yahya Momtaz, Stefania Conte, Guido Russo</i>	156
Digital Explorations of Historical Multilingual Practices. The Challenges of the HyperAzpilcueta Project <i>Manuela Bragagnolo, Marcus Pöckelmann, Polina Solonets, Andreas Wagner</i>	160
Digitalizzazione di un fondo archivistico per la creazione di un centro di documentazione digitale <i>Dario Baldini</i>	165
Risorgimento Digitale: Un progetto di hyperedizione per i testi risorgimentali. Le Noterelle di Abba come caso di studio <i>Vincent Mobilia</i>	171
SpaceLat: La geografia della letteratura latina tardoantica <i>Riccardo Consolini</i>	177
Verso l'edizione digitale del carteggio Canneti-Fiacchi <i>Chiara Manca, Fiammetta Sabba, Bianca Sorbara, Silvia Tripodi</i>	184
L'edizione critica digitale della 'Scienza nuova' di Giambattista Vico in Scholarly Digital Edition <i>Alessia Scognamiglio, Roberto Evangelista, Manuela Sanna, Salvatore Prinzi, Stefano Veneroni, Chiara Aiola, Luca de Santis</i>	190
L'edizione digitale del papiro P.Tor.Choach. 12 in collaborazione con il Museo Egizio di Torino <i>Chiara Senatore</i>	197
Dal palcoscenico al digitale: modelli di data visualization per la valorizzazione dell'Archivio Teatro delle Marionette di Gianni e Cosetta Colla <i>Elena Radaelli</i>	202
Metodologie computazionali per l'organizzazione di archivi nati digitalmente <i>Mariangela Giglio</i>	208
Archivi digitali per la conservazione e valorizzazione del patrimonio culturale: il caso del Santuario della Madonna di Carufo <i>Caterina Ciccotti</i>	215

Il progetto ArPeR. Per un Archivio dei periodici romaneschi <i>Martina Ludovisi</i>	222
Un modello integrato per il Roman d'Alexandre del codice Correr 1493: annotazione linguistica e edizione critica digitale <i>Giacomo Costa, Simone Zenzaro</i>	228
Descrivere la catastrofe: documentare la diegesi per la catalogazione di opere distopiche e post-apocalittiche <i>Luca Paolo Bruno, Valeria Stabile, Juan Scassa, Carmelo Caruso, Ludovica Pannitto</i>	232
Un'applicazione pratica per l'edizione digitale di testi agiografici e calendariali <i>Luca Avellis</i>	238
Embracing flexibility: new EVT features for critical editing, accessibility and inclusivity <i>Roberto Rosselli Del Turco, Davide Cucurnia, Marina Buzzoni</i>	244
Riscoperte poliane: edizione digitale di un manoscritto inedito de Il Milione <i>Giulia Fabbris, Samuela Simion, Fabio Soncin</i>	251
«Proximior perfectioni»: criticità e future prospettive del progetto Dante Limina <i>Elisabetta Tonello</i>	257
TEI Encoding as a Unified Structure for Multilingual Digital Editions: The LeggoManzoni Case Study <i>Mariia Levchenko, Beatrice Nava, Ersilia Russo</i>	264
Verso l'implementazione di un sistema di riconoscimento di allusioni al lessico dantesco nelle testimonianze del Lager: il caso d'uso in Voci dall'Inferno <i>Carla Congiu, Angelo Mario Del Grosso, Marina Riccucci</i>	270
Il progetto RETI (REndering Texts and Images): metodologia e primi risultati <i>Chiara Barbero, Matteo Di Franco, Federica Lazzerini, Annamaria Persia</i>	276
Edizione digitale ed autorialità plurima: quali sfide? <i>Stefania Tesser</i>	283
Il corpus di prosa letteraria del progetto RIND (1830-1930). Assunti teorici e vincoli pratici <i>Stefano Ondelli, Pietro Mazzarisi</i>	289
ZoneRW: verso un'integrazione con Kraken ed eScriptorium per il riconoscimento e la gestione avanzata delle regioni di interesse <i>Pietro Sichera, Angelo Mario Del Grosso, Laura Mazzagufu, Daria Spampinato</i>	297
L'edizione digitale dei primi 16 Taccuini di Paolo Orsi <i>Giuseppina Monterosso, Andrea Bolioli, Elisa Bonacini, Gianmario Cattaneo, Dario Gonella, Anna Maria Marras, Salvatore Spina, Paola Venuti</i>	303
Artificial intelligence vs human handwriting: annotating damaged manuscripts <i>Dumitru Scutelnic, Laura Gazzani, Paolo Pellegrini, Claudia Daffara</i>	308
Modellazione, interoperabilità e riuso in DiScEPT <i>Tiziana Mancinelli, Hansmichael Hohenegger, Federico Boschetti, Angelo Mario Del Grosso, Eleonora De Longis, Gloria Mugelli</i>	314
Navigating the Digital Transition: Lessons from a Hybrid Critical Edition Project <i>Elisa Bastianello, Reto Baumgartner</i>	319
Human-LLM Synergy in Higher Education Publishing: Two ChatGPT Use Cases within Editorial Pipelines <i>Gianluca Pavani</i>	327
Taming the Hydra: A Model for Textual Dynamics and Constellations of Goethe's Venetian Epigrams <i>Daniele Fusi, Matteo Zupancic, Franz Fischer, Claus Zittel</i>	334

Computational Methods

The Influence of AI Tools on University Students' Writing Style: A Stylometric Analysis of Narrative Texts <i>Dimitris Bilianos, Katerina Florou</i>	343
Usare i Large Language Model per l'analisi del testo narrativo: strategie di prompt engineering per il riconoscimento del discorso indiretto libero nella narrativa italiana 1830-1930 <i>Aurora Argenzio, Fabio Ciotti, Anna Chiara Corradino</i>	349
Historical GIS e metodologie digitali per una storia della copertura boschiva <i>Vincenzo Colaprice</i>	357
Experiments on the Use of LLMs for the Translation of the Babylonian Talmud <i>Mafalda Papini, Davide Albanesi, David Dattilo, Emiliano Giovannetti, Simone Marchi</i>	363
Metodi di allineamento testuale bilingue per un'edizione genetica digitale dei Mémoires di Carlo Goldoni <i>Matteo Zibardi</i>	368
Eastern Law in Western Words: Analyzing Roman Legal Terminology in Medieval Charters <i>Tamás Kovács, Angelos Nikolaou, Johannes Laroche, Georg Vogeler</i>	375
Il corpus del Digesto: approcci e metodi computazionali per la creazione di risorse linguistiche <i>Alessandra Cinini, Paola Marongiu, Eva Sassolini</i>	379
Preliminary Results for the Explanation of Neural Network-based Handwriting Identification in Historical Manuscripts <i>Riccardo De Cesaris, Valerio Caravani, Arianna Pastorini, Serena Ammirati, Paolo Meriardo</i>	386
From Documents to Data: Digital Technologies in the Study of Notarial Charters <i>Franziska Decker, Sandy Aoun, Giuseppe Consolo</i>	392
Verso la svolta computazionale della critica dantesca <i>Fara Autiero, Vittorio Celotto, Gennaro Ferrante, Chiara Fusco, Sandra Gorla, Giuseppe Andrea Liberti, Mariangela Palomba, Serena Picarelli, Stefano Angelo Rizzo, Silvia Tripodi</i>	397
Phylo-1-preview. Un modello T5-Base per l'emendazione dei testi antichi <i>Giuseppe Ferrara</i>	404
«Glottolab: A Linguistic Adventure»: Lo sviluppo di un'attività gamificata incentrata sulla linguistica <i>Cecilia Cattaneo, Claudia Roberta Combei, Chiara Zanchi</i>	411
Concordanze e NLP: idee, metodi e regole per l'applicazione alla lingua italiana <i>Pietro Sichera, Christian D'Agata, Giuseppe Palazzolo</i>	419
Reverse Engineering Critical Apparatuses for HTR Ground Truth Creation: The Case of Kennicott's Collation of the Hebrew Bible <i>Luigi Bambaci, Nachum Dershowitz, Daniel Stökl Ben Ezra</i>	426

Data and Knowledge Representation

Prototyping an Atlas of Early Modern English Drama: An Experiment on DraCor Data <i>Luca Giovannini, Andreas Wagner</i>	435
ATLAS: A data model for describing FAIR Digital Humanities research outcomes <i>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</i>	440

Cantautorato e Digital Humanities. Per una valorizzazione dell'opera di Fabrizio De André, Lucio Dalla, Gianmaria Testa <i>Marcello Ranieri</i>	448
OWL Ontology on the European Integration Process between 1949 and 1979 <i>Lorenzo Galvagno</i>	453
LiITA, una Knowledge Base di risorse interconnesse per l'italiano <i>Eleonora Litta, Marco Passarotti, Paolo Brasolin, Valerio Basile, Cristina Bosco, Andrea Di Fabio</i>	460
Dai limina a LiMINA: un database per i marginalia alla Commedia <i>Serena Malatesta, Beatrice Mosca</i>	466
Dai Materiali Didattici alle Piattaforme FAIR: Costruire un'Infrastruttura di Training in H2IOSC <i>Giulia Pedonese, Francesca Frontini, Roberta Ottaviani, Federico Boschetti, Alessia Spadi, Lucia Francalanci, Alessia Scognamiglio, Pietro Restaneo, Antonina Chaban, Jana Striova, Laura Benassi</i>	473
IlluminAI: un sistema di navigazione interattivo per i manoscritti miniati rinascimentali <i>Valeria Minisini, Giorgio Gosti, Bruno Fanini</i>	478
Making Germanic Cultural Heritage accessible to students: a proposal for a case study <i>Chiara De Bastiani, Giulia Fabbris</i>	485
/DH.arc Vocabularies: Making semantic artefacts more visible and accessible using SKOS <i>Laurent Fintoni</i>	492
Modeling an Ontology for Heritage Science: Challenges and Key Strategies <i>Erica Scarpa, Riccardo Valente, Irene Rossi</i>	499
Linked Open Data and IIIF for connecting manuscripts images with their transcriptions: a case study from the Veneranda Biblioteca Ambrosiana <i>Lorenza Talarico</i>	505
Describing Monastic Iconography Using Semantic Data: A Preliminary Investigation <i>Sofia Baroncini, Francesco Mele</i>	511
A Linguistic Knowledge Graph of Word Borrowings from Portuguese <i>Anas Fahad Khan, Ana Salgado</i>	519
E.T and Visual Culture Ontology (ETVCO): Perspectives on Extraterrestrial Influence in Visual Heritage <i>Kaosaier Wusiman, Simone Casazza</i>	524
Automating XML-TEI Encoding of Unpublished Correspondence: A Comparative Analysis of two LLM Approaches <i>Marco De Cristofaro, Daniel Zilio</i>	531
Modelli e tecnologie integrate e innovative per una cittadinanza digitale equa e sostenibile <i>Cristina Marras, Vittoria Fabiani, Enrico Pasini, Lisa Reggiani, Pietro Sichera, Paolo Ongaro, Martina Rossi</i>	538
A Case Study in Cultural Heritage: A System Linking Three Open Data Tools – Digital Philology for Dummies (DPHD), Edition Visualization Technology (EVT), and a Relational Database <i>Renato Caenaro, Chantal Pivetta</i>	544
Modeling Intermediality and Interpretations in Contemporary Combinatory Literature: Revealing Il Giuoco dell'Oca by Edoardo Sanguineti <i>Enrica Bruno, Maria Francesca Bocchi, Francesca Tomasi</i>	551
From Metadata to Storytelling: A Framework For 3D Cultural Heritage Visualization on RDF Data <i>Sebastian Barzaghi, Simona Colitti, Arianna Moretti, Giulia Renda</i>	558
Between Text and Icon: Towards A Representational Model for Ekphrastic Relations <i>Maria Francesca Bocchi, Carlo Teo Pedretti, Fabio Vitali</i>	566

Preservation of Memory and Digital Cultural Heritage

Life and Death of DH Projects: A Preliminary Investigation of Their Lifecycles in Italy <i>Erica Andreose, Giorgia Crosilla, Remo Grillo, Gianmarco Spinaci</i>	575
Research on Street Art in the Digital Space <i>Aleksandra Tselikova</i>	581
Motion Visualisation of Dancers' Performances <i>Giacomo Alliaia, Loïc Serafin, Samy Mannane, Sarah Kenderdine</i>	587
Entità in relazione: policies, soluzioni tecnologiche e modelli lessicali per un (eco)sistema informativo integrato <i>Herbert Natta, Michela Tardella, Eleonora Lattanzi, Gianluca Rossi, Roberta Maggi</i>	593
Preserving and enhancing cultural heritage: the Digest project <i>Alessandra Cinini, Paola Marongiu, Eva Sassolini, Monica Monachini</i>	600
The Staccioli Digital Archive: Using Knowledge Graphs to power digital art history catalogues and art exhibitions <i>Klaus Werner, Pietro Liuzzo, Alessandro Adamou</i>	608
Fantàsimè: Interactive Drama per la valorizzazione del Patrimonio Culturale <i>Maria Chiara Provenzano, Eleonora Miccoli, Mario A. Bochicchio</i>	613
Soluzioni phygital e mediazione culturale: riflessioni digiteconomiche nell'era dell'IA <i>Nicola Barbuti, Mauro De Bari</i>	619
MeMo: Una mappa letteraria digitale per la memoria del Mezzogiorno <i>Laura Giurdanella, Giuseppe Palazzolo, Bernardo De Luca, Fara Autiero, Marco Gatto, Sabatino Peluso, Concetta Maria Pagliuca, Andrea Schembari</i>	627
The relationship between art and sound: An experiment on the engagement of the cultural tourist <i>Sara Benetti, Nicola Orio</i>	633
Analisi RTI delle iscrizioni runiche del Leone del Pireo (Arsenale di Venezia) <i>Paola Peratello, Elisa Corrà</i>	639
Torino anni Ottanta. Digitalizzazione del patrimonio documentario e ricostruzione virtuale delle mostre negli spazi pubblici e privati <i>Filippo Yahia Masri</i>	646
Linguistica dei corpora e informatica umanistica per la valorizzazione plurilingue del patrimonio culturale: implementazione del progetto UniVOCIttà <i>Rita Gramellini, Valeria Zotti</i>	651
«Il mio sommario dunque è tutto qui?» Per Franco Fortini <i>Emmanuela Carbé, Mariangela Giglio, Pietro Orlandi, Jacopo Maria Romano, Giulio Quaresima</i>	658

Digital Explorations of Historical Multilingual Practices. The Challenges of the HyperAzpilcueta Project

Manuela Bragagnolo¹, Marcus Pöckelmann², Polina Solonets³, Andreas Wagner⁴

¹ University of Trento, Italy – manuela.bragagnolo@unitn.it

² Martin Luther University Halle-Wittenberg, Germany – marcus.poeckelmann@informatik.uni-halle.de

³ Max Planck Institute for Legal History and Legal Theory, Germany – solonets@lhlt.mpg.de

⁴ Max Planck Institute for Legal History and Legal Theory, Germany – wagner@lhlt.mpg.de

ABSTRACT (ENGLISH)

Based on the example of an early modern case of self-translation and collaborative translation, the contribution seeks to open discussion about the relation between multilingual practices embodied in textual research objects on the one side and the “user experience” of digital platforms and their affordances on the other.

Keywords: text collation; digital multilingual practices; HyperAzpilcueta

ABSTRACT (ITALIANO)

Esplorazioni digitali di pratiche multi linguistiche storiche. Le sfide del progetto HyperAzpilcueta

Basandosi sull'esempio di un caso di autotraduzione e traduzione collaborativa della prima età moderna, il contributo cerca di aprire una discussione sulla relazione tra le pratiche multilinguistiche incarnate negli oggetti di ricerca testuale, da un lato, e la “user experience” delle piattaforme digitali e delle loro possibilità, dall'altro.

Parole chiave: collazione di testi; pratiche digitali multilinguistiche; HyperAzpilcueta

1. INTRODUCTION

Historical multilingual practices involve much more than mere linguistic translation. For this reason, conceptualizing the main aspects of those practices and their application in different textual witnesses is a challenge in itself. But the question then is: how can digital environments and methods facilitate investigation of the various aspects of these practices and their differences?

As an illustration of the general problem, we present “HyperAzpilcueta”, a digital synoptic edition of several printed editions of an early modern bestseller, *Manual de Confessores* (Manual for Confessors). Its author, the Spanish canon law professor Martín de Azpilcueta (1492-1586), translated and simultaneously substantially revised his work several times, incorporating different types of authorities, examples and arguments in the different editions. Together with his printers, he even organized typographical devices for his respective target audience to better orient themselves in the text. With the help of translations and revisions, he thus devised new ways of producing different types of knowledge of (religious, spiritual, but also legal) normativity.

2. MARTÍN DE AZPILCUETA’S MANUAL DE CONFESORES AS A CASE OF MULTILINGUAL PRACTICE

First printed in Portugal in 1549, the *Manual* was then transformed, updated, and translated into different languages by Azpilcueta himself, in a revision process that lasted for almost 40 years. Four of the editions showcasing the most important transformations are used in the HyperAzpilcueta project: the first edition, printed in Coimbra in Portuguese in 1549; the second Portuguese edition, with major transformations (Coimbra, Portuguese, 1552); the 1556 Spanish translation by Azpilcueta, largely enriched (Salamanca, 1556); and finally, the Latin translation, completely revised in structure (Rome, 1573).

As Azpilcueta was working on translation alone or in cooperation with other scholars and practitioners of the time, his work on the *Manual* is an example of both, “collaborative translation” (Cordingley & Frigau Manning, 2017) and “self-translation” practices (Miglietti, 2024). Such a multilingual practice was quite common in the early modern world, with every new edition used to update, restructure, and adapt the text (and the book as a material object) to new target audiences, new contexts of production and circulation, and to changes in societal practices that form the subject matter of many parts of the book.

In what dimensions can we study the translation practice of the *Manual*?

Firstly, the differences between the texts reflect the varying contexts of writing, book production and dissemination. They actually “accompanied” Azpilcueta’s moving from Portugal (Coimbra, 1549; Coimbra,

1552) to Spain (Salamanca, 1556), and from Spain to Rome (Rome, 1573). Every new edition and “self-” or “collaborative” translation was produced in close contact with a different set of local actors (Bragagnolo, 2022; Bragagnolo, 2024): throughout his career, missionaries and confessors have contacted Azpilcueta with questions and doubts about the latest (or earlier) edition(s) of the *Manual*, as did kings and penitents, theologians and jurists, not to forget book printers and traders. Economic and political constellations varied locally and changed substantially during the 40 years of revising the *Manual*, and the Council of Trent (1545–1563) revised in important ways the official Catholic doctrine on many questions treated in it. Therefore, each new version incorporated the feedback of a particular community of practice in its revised and translated language.

Secondly, the different versions of the *Manual* aimed to reach different reader groups: The first edition (in 8° pragmatic vernacular handbook) was meant for simple and unlearned confessors, active in remote parishes of Portuguese provinces. By contrast, the very erudite, rich, in 4° Latin *Enchiridion* was intended to reach learned confessors and missionaries of the universal Church, widespread all over the world, via academic channels of communication and book distribution. The editions in between reflected this path in their language, from the vernacular to Latin, with injections of doctrine, and increasingly technical, complex and erudite use of language.

Those differences between contexts of production and target audiences found expression in concrete linguistic phenomena: different audiences demand different sets of references to authorities, different contexts provide different sets of examples (mundane or contrived, colonial or courtly), different languages have different proverbs and personae. And in terms of knowledge organization, Azpilcueta developed different ways of mapping and navigating the subject matter: even though the basic structure of 27 chapters was kept intact in all editions, each edition shows another ordering of arguments, examples and references, mostly with new sections added. Different structural and typographical finding devices such as sub-chapter headings and summaria, the wrapping of newly added passages in asterisks or their introduction by daggers etc. suggest very different uses of the book by the readers. Some of them, such as the marginal numbers, even seem to indicate an (again, multilingual) practice of navigating across and comparing the different editions on the part of the readers.

3. CONCEPTUALIZING THE COMPREHENSION OF MULTILINGUAL PRACTICE IN A DIGITAL PLATFORM

The question arises of how a digital platform can facilitate in the study of the various described aspects of the *Manual* and texts with a similar history of tradition. The following section discusses the main requirements and the possibilities already offered by existing approaches.

First, it is essential to allow researchers to closely examine the text itself by presenting it in full detail whenever necessary (close-reading). This should also take into account – may even by optional highlighting – their various special elements such as different fonts, marginal numbers, literature references or special characters (e.g. asterisks and daggers in our case). If the texts are encoded in TEI XML, this is supported by many tools for generating digital editions such as TEI-Publisher,¹ EVT² (Rosselli Del Turco et al., 2014) or the Versioning Machine³ (Schreibman et al., 2003).

A second essential feature is the possibility to compare certain passages of the text witnesses with each other in order to allow researchers to identify similarities and textual variants. Typical close-reading representations for this purpose in collation tools highlight or arrange the words in specific ways. This includes classic synoptic views, which can be found in the already mentioned tools (TEI-Publisher, EVT and Versioning Machine) as well as in LERA (Pöckelmann et al., 2022), or variant graphs as in CollateX (Dekker et al., 2015) or TRAViz (Jänicke et al., 2015). However, the underlying collation algorithms in current tools are typically designed for witnesses of the same language and rather simple textual variants like the replacement of individual words or phrases with alternatives. While research for handling multiple languages in the domain of scholarly editions is presently going on (Balck et al., 2025; Reboul, 2024; Hohenegger et al., 2023), available collation tools still lack the ability to automatically point the researcher to semantic similarities and differences on a word or phrase level when it comes to multilingual texts. When dealing with larger amounts of text like the *Manual* has, the collation in hierarchical steps is reasonable (Pöckelmann et al., 2022), i.e., in a first step larger segments of the text such as sentences or

¹ <https://teipublisher.com/exist/apps/tei-publisher-home/index.html> (cons. 05.01.2025).

² <http://evt.labcd.unipi.it> (cons. 05.01.2025).

³ <http://v-machine.org> (cons. 05.01.2025).

paragraphs are aligned before in a second step the collation on the word level is performed. This presupposes the constitution and alignment of those segments in the first place, which becomes harder the more substantial the revisions of the text witnesses are. Besides the presentation and user interface, features of automatic segmentation and detection of semantic similar phrases or sentences that differ only on the surface, i.e., mostly in a changed wording but not meaning, are relevant for the platforms or toolchains we are interested in. Promising recent solutions outside the domain of scholarly editing are based on word or sentence embeddings (e.g. Kusner et al., 2015). While these approaches have been intended for monolingual applications at first, they were soon be extended to the multilingual realm (Reimers & Gurevych, 2020; Ham & Kim, 2021), resulting in state-of-art approaches for effective sentence-alignment even across languages, e.g. SentAlign (Steingrímsson et al., 2023). However, these have difficulties when dealing with major structural differences among the text witnesses, as it is the case with the *Manual*. Another problem lies in the training of the underlying embeddings, which requires a large amount of digitized text. This is simply not available as we face so-called *low-resource languages* in this project due to the historical state and the specific genre. In our alternative approach to embeddings, we apply a rule-based system for segmentation with an alignment algorithm that profits from the similarity of the languages in the *Manual* (Wagner & Bragagnolo, 2019) in combination with an evaluation of some of the particular features mentioned above and are currently extending the collation platform LERA⁴ (Pöckelmann et al., 2022) accordingly.

Even if the used algorithms deliver good results, it is unlikely that any automatic segmentation or alignment will be perfect at extracting and evaluating all the relevant features nor that the results correspond exactly to the researcher's expectations. This weakness is amplified by the fact that often changes in one aspect (like segmentation) entail substantial further changes in another (alignment). Thus, a digital platform must provide means for researchers to correct both segmentation and alignment in the later close reading process. It should be possible to join or split segments, move them around (in the sense of adjusting their position in the synoptic view), and even place copies of existing segments – marked as such – at other positions to reflect transpositions.

Finally, the question arises as to how the entirety of the work and the identified revisions can be presented. With the large volume of text on the one hand and the structural changes in terms of both sequencing and volume on the other, there is a need for overview visualizations (distant-reading) in order to navigate through these volumes of text as well as to compare structural changes. The latter may include transpositions that span over long distances within the different text witnesses. A recent survey (Yousef & Jänicke, 2021) lists a huge variety of visualization methods that has been developed over the years to represent alignments, some of which are applicable to huge text witnesses for distant-reading. Furthermore, some older approaches have focused on conveying the dynamics of revision processes or structural aspects (e.g. Fry, 2009; Cheesman & Roos, 2017). However, due to the combination of various aspects of the *Manual* – multilinguality, restructuring and the mere existence of more than two witnesses – no existing visualization is sufficient. A flexible, zoomable, or elastic way of providing overviews of different levels of distance seems desirable (as in Bludau et al., 2020; Pöckelmann et al., 2024). This could include the text itself at the lowest zoom level, making it a scalable-reading approach.

4. CONCLUSION

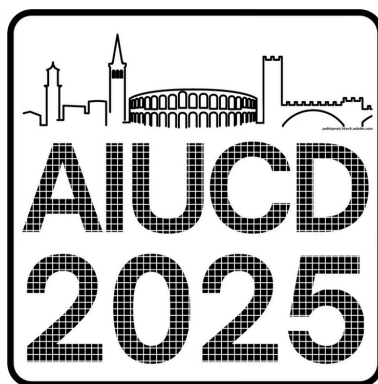
As we have shown, for many of the mentioned requirements, some functionalities are, in principle, available in existing tools. Yet the combination of tasks (represent structural as well as semantic changes, establish alignment, facilitate close reading, process etc.), the volume of text to be processed and presented and the multitude of features to take into account – and the multilinguality affecting all the tasks individually and in combination – mean that there is, as of now, no platform that really does facilitate study of historical multilingual practices like we have presented with Apilcueta's *Manual*. Moreover, to the best of our knowledge, for the issue of distinguishing "salient" from "trivial" differences between texts in different languages (where practically everything is different in terms of linguistic surface forms), no solution at all is available. The state of affairs rather seems to demand a discussion about how to conceptualize "salient semantic similarity or difference" in a multilingual context, how to anchor any mechanism in this regard to concrete words or phrases in the source documents, and (only then) how to convey eventually acquired measurements in a user interface.

⁴ <https://lera.uzi.uni-halle.de> (cons. 05.01.2025).

REFERENCES

- Balck, S., Dähne, J., Etling, F., Fischer, F., Frenzel, S., Grote, B., Heße, S., Molitor, P., Pöckelmann, M., Ritter, J., Singh, Y. & Stede, M. (2025). *Collation of Multilingual Versions of a Text: Necessity, Approach, Challenges* [Conference presentation abstract]. ADHO DH2025. Lisbon; Portugal.
- Bludau, M.-J., Brüggemann, V., Busch, A. & Dörk, M. (2020). Reading Traces: Scalable Exploration in Elastic Visualizations of Cultural Heritage Data. *Computer Graphics Forum*, 39(3), 77-87. <https://doi.org/10.1111/cgf.13964>.
- Bragagnolo, M. (2022). Les voyages du droit du Portugal à Rome. Le Manual de confesseurs de Martín de Azpilcueta (1492-1586) et ses traductions. In J.-L. Fournel, I. Paccagnella (Eds.), *Traduire. Tradurre. Translating. Vie des mots et voies des œuvres dans l'Europe de la Renaissance* (pp. 209-228), Droz.
- Bragagnolo, M. (Ed.). (2024). *The Production of Knowledge of Normativity in the Age of the Printing Press. Martín de Azpilcueta's Manual de Confessores in global perspective*, Brill. <https://doi.org/10.1163/9789004687042>
- Cheesman, T. & Roos, A., (2017). Version Variation Visualization (VVV): Case Studies on the Hebrew Haggadah in English. In M. Büchler & L. Mellerin (Eds.), *Computer-Aided Processing of Intertextuality in Ancient Languages* [Special Issue]. *Journal of Data Mining & Digital Humanities*. <https://doi.org/10.46298/jdmdh.3704>
- Cordingley, A. & Frigau Manning, C. (Eds.) (2017). *Collaborative Translation: from the Renaissance to the Digital Age*, Bloomsbury Academic. <https://doi.org/10.5040/9781350006034>
- Dekker, R., Van Hulle, D., Middell, G., Neyt, V. & Van Zundert, J. (2015). Computer-supported collation of modern manuscripts: CollateX and the Beckett Digital Manuscript Project. *Digital Scholarship in the Humanities*, 30(3), 452-470. <https://doi.org/10.1093/llc/fqu007>
- Fry, B. (2009) *On the Origin of Species: The Preservation of Favoured Traces*. <https://benfry.com/traces/>
- Ham, J., Kim, E.-S. (2021). Semantic alignment with calibrated similarity for multilingual sentence embedding. In M.-F. Moens, X. Huang, L. Specia & S. W. Yih (Eds.), *Findings of the Association for Computational Linguistics: EMNLP 2021* (pp. 1781-1791). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.findings-emnlp.153>
- Fry, B. (2009) *On the Origin of Species: The Preservation of Favoured Traces*. <https://benfry.com/traces/>
- Hohenegger, H., Montemagni, S., Pasini, E., Ciotti, F., Mancinelli, T., De Longis, E., Chiaravalli, I., Boschetti, F., Del Grosso, A. M. & Meschini, F. (2023) *DiScEPT – Digital Scholarly Editions Platform and aligned Translations*. <https://www.studigermanici.it/progetto-discept>
- Jänicke, S., Geßner, A., Franzini, G., Terras, M., Mahony, S. & Scheuermann, G. (2015). TRAViz: A Visualization for Variant Graphs. *Digital Scholarship in the Humanities*, 30(suppl_1), i83–i99. <https://doi.org/10.1093/llc/fqv049>
- Kusner, M. J., Sun, Y., Kolkin, N. I. & Weinberger, K. Q. (2015). From Word Embeddings to Document Distances. In F. Bach & D. Blei (Eds.), *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*. (pp. 957-966). <https://dl.acm.org/doi/10.5555/3045118.3045221>
- Miglietti, S. (2024). Translations of self-translations in early modern Europe: preliminary remarks and exploratory case studies. *Philological Quarterly*, 102(2-3), 199-223.
- Pöckelmann, M., Damm, A. & Rebigier, B. (2024). *An Interactive Visualization of Recurring Microforms Across Various Hebrew Works and their Witnesses* [Conference presentation abstract]. DH2024, 34. international annual conference of Digital Humanities, Washington D.C., United States.
- Pöckelmann, M., Medek, A., Ritter, J. & Molitor, P. (2022). LERA—an interactive platform for synoptical representations of multiple text witnesses. *Digital Scholarship in the Humanities*, 38(1), 330–346. <https://doi.org/10.1093/llc/fqac021>
- Reboul, M. (2024). Alignement sémantique et manque de données : l'apport des modèles de langue. Le cas du latin et du grec. In B. Bordalejo, R. Risam & E. Château-Dutier (Eds.), *DH Unbound 2022, Selected Papers* [Special Issue]. *Digital Studies/Le champ numérique* 13(3): 1–13. <https://doi.org/10.16995/dscn.9669>.
- Reimers, N. & Gurevych, I. (2020). Making monolingual sentence embeddings multilingual using knowledge distillation. In B. Webber, T. Cohn, Y. He & Y. Liu. (Eds.), *Proceedings of the 2020*

- Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 4512-4525). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.emnlp-main.365>
- Rosselli Del Turco, R., Buomprisco, G., Di Pietro, C., Kenny, J., Masotti, R. & Pugliese, J. (2014). Edition Visualization Technology: A Simple Tool to Visualize TEI-based Digital Editions. *Journal of the Text Encoding Initiative*, 8. <https://doi.org/10.4000/jtei.1077>
- Schreibman, S., Kumar, A. & McDonald, J. (2003). The versioning machine. *Literary and Linguistic Computing*, 18(1), 101–107. <https://doi.org/10.1093/lc/18.1.101>
- Steingrímsson, S., Loftsson, H. & Way, A. (2023). SentAlign: Accurate and scalable sentence alignment. In Y. Feng & E. Lefever (Eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (pp. 256-263) <https://doi.org/10.18653/v1/2023.emnlp-demo.22>
- Wagner, A. & Bragagnolo, M. (2019). *Multimodale Versuche der Alignierung historischer Texte*. [Conference presentation abstract] DHd2019, 6. Jahrestagung der Digital Humanities im deutschsprachigen Raum. Frankfurt, Germany, 181–184. <https://doi.org/10.5281/zenodo.4622224>.
- Yousef, Y. & Jänicke, S. (2021). A Survey of Text Alignment Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 27(2), 1149–1159. <https://doi.org/10.1109/TVCG.2020.3028975>



**UNIVERSITÀ
di VERONA**
Dipartimento
di LINGUE
E LETTERATURE STRANIERE



ASSOCIAZIONE per
l'INFORMATICA UMANISTICA
e la CULTURA DIGITALE

