European Association for Digital Humanities 2026

"Linking Europe: Digital Humanities Without Borders"

Book of Abstracts

Jagiellonian University, Kraków, Poland
15-19 September 2026

Under the Honorary Patronage of the Rector of the Jagiellonian University

eadh EUROPEAN ASSOCIATION FOR DIGITAL HUMANITIES

JAGIELLONIAN UNIVERSITY IN KRAKÓW

PAN IJP Instytut Języka Polskiego Polskiej Akademii Nauk

Uniwersytet Wrocławski

JAGIELLOŃSKIE CENTRUM HUMANISTYKI CYFROWEJ

Digital Humanities Lab
Flagship Project at Jagiellonian University

RESEARCH UNIVERSITY
EXCELLENCE INITIATIVE

Book of Abstracts

European Association for Digital Humanities 2026
"Linking Europe: Digital Humanities Without Borders"

15-19 September 2026, Jagiellonian University, Kraków, Poland

Co-organised by the Institute of Polish Language of the Polish Academy of Sciences and the University of Wrocław

Layout and typesetting
Justyna Łukaszyk
Logo design
Patrycja Wojkowska
Edited by
Aleksandra Rykowska
Proofreading
Aleksandra Rykowska

ISBN 978-83-977695-1-9
DOI 22078032

Table of contents

AI Statement

In the preparation of this abstract, ChatGPT (OpenAI) was used to assist in rephrasing and restructuring portions of the text in order to improve readability, clarity, and overall flow. The tool was employed solely for linguistic refinement and stylistic optimization. All substantive content, technical descriptions, research framing, and scholarly claims originate from the authors. The authors reviewed and revised all AI-assisted output and take full responsibility for the final version of the text.

References

Abitbol, Roy, Ilan Shimshoni, and Jonathan Ben-Dov. 2021. "Machine Learning Based Assembly of Fragments of Ancient Papyrus." *Journal on Computing and Cultural Heritage* 14 (3): 1–21. https://doi.org/10.1145/3460961.

Ostertag, Cecilia, and Marie Beurton-Aimar. 2021. "Using Graph Neural Networks to Reconstruct Ancient Documents." In *Pattern Recognition. ICPR International Workshops and Challenges*, edited by Alberto Del Bimbo, Rita Cucchiara, Stan Sclaroff, et al., vol. 12667. Springer International Publishing. https://doi.org/10.1007/978-3-030-68787-8_3.

Pirrone, Antoine, Marie Beurton Aimar, and Nicholas Journet. 2019. "Papy-S-Net: A Siamese Network to Match Papyrus Fragments." *Proceedings of the 5th International Workshop on Historical Document Imaging and Processing* (Sydney NSW Australia), September, 78–83. https://doi.org/10.1145/3352631.3352646.

Pirrone, Antoine, Marie Beurton-Aimar, and Nicholas Journet. 2021. "Self-Supervised Deep Metric Learning for Ancient Papyrus Fragments Retrieval." *International Journal on Document Analysis and Recognition (IJDAR)* 24 (3): 219–34. https://doi.org/10.1007/s10032-021-00369-1.

Vu, Manh Tu, and Marie Beurton-Aimar. 2024. "ViT-ED: Transformer Network for Image Similarity Measurement." In *Document Analysis and Recognition – ICDAR 2024*, edited by Elisa H. Barney Smith, Marcus Liwicki, and Liangrui Peng. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-70546-5_18.

POSTER

Modelling Early Modern Reasoning for the Venetian Risposte: a Scalable Framework

Emmanuela Carbé Ca' Foscari University of Venice – VeDPH
Federico Boschetti ILC CNR Pisa – VeDPH, Italy
Isabella Cecchini ISEM CNR Rome, Italy
emmanuela.carbe@unive.it, federico.boschetti@ilc.cnr.it

Policy reasoning, Data Modelling, Early modern Venice, Archival series, HTR

Abstract

The documentary series *Risposte* (replies), preserved in the Archivio di Stato di Venezia, forms an almost continuous collection of advisory opinions drafted by the Cinque Savi alla Mercanzia for the Venetian Senate (1540-1733). The paper presents a pilot study that treats the *Risposte* as structured units of policy reasoning that can be modelled and analysed at scale. Focusing on register 142 (1607-1610), for which a complete digitisation and a subset of manual diplomatic transcriptions are available, we define the single *Risposta* as the primary decision unit and design a two-layer data model: a minimal documentary core plus an analytical layer with controlled vocabularies (policy domain, decision orientation, triggers, actors, geography). Institutional variants are encoded as explicit relations rather than treated as exceptions. To support scalability, we report a first evaluation of an automated text recognition (ATR) workflow bridging Transkribus and eScriptorium through the XML-PAGE interchange format, and we prototype lightweight rule-based procedures (deontic formula extraction, monetary expressions, similarity measures) for semi-automatic population and quality control. The pilot operationalizes two directions of scalability: a stable decision-unit model and an interoperable transcription pipeline to populate it, enabling future large-scale

analysis of institutional coherence, policy drift, and trade oversight in early modern Venice.

Introduction

The documentary series *Risposte* (replies), preserved in the Archivio di Stato di Venezia, constitutes an almost continuous collection of advisory opinions drafted by the Cinque Savi alla Mercanzia for the Venetian Senate between 1540 and 1733 (Borgherini-Scarabellin 1925). Established as a permanent magistracy in 1517, the Cinque Savi operated at the intersection of mercantile practice and governmental decision-making, addressing matters of trade regulation, citizenship privileges, consular tariffs, monetary stability, and fiscal enforcement.

The period between approximately 1570 and 1670 witnessed profound geopolitical transformations that reshaped Mediterranean trade networks and challenged Venice's commercial position. Economic historians have long drawn on the *Risposte* to document these changes (Fusaro 2015; Pezzolo 2013; Sella 1968, 1994); Borgherini-Scarabellin (1925) remains the only monograph devoted to the magistracy, and the series itself still awaits a dedicated modern study and representation as structured data. Rather than treating the *Risposte* solely as narrative evidence of economic change, this project approaches them as structured units of policy reasoning and investigates whether the series can be modelled as a coherent governance system (de Vivo 2007).

This defines a specific data-modelling problem (Flanders & Jannidis 2019). Recent computational work on Venetian serial sources operates on records whose structure the source itself supplies: the 1808 cadastre pairs every parcel with fixed fields for number, owner, function and area (di Lenardo et al. 2025; Karch et al. 2025), and the *Garzoni* corpus of some 55,000 apprenticeship contracts records master, apprentice, trade and term in equally regular slots (Colavizza 2017). A *Risposta* offers far less: a date, a marginal note, the signatures of the Savi, and between them an argued opinion. The note is a scribal finding aid, not a classification: it points to a subject, while the decision itself, its orientation and its conditions, exists only in the prose, in formulas such as *stimiamo* (we deem), *debba* (must) or *non possa* (may not) and in the argument built around them. That problem is the difference between structure supplied by the source and structure read into it.

Proof of Concept and Data Model

The proof of concept is built on register 142 (1607-1610), for which a complete photographic digitisation and a subset of manual diplomatic transcriptions are available. On this material the pilot tests two complementary directions of scalability: model scalability, through a stable decision-unit model with a minimal schema, and quantity scalability, through an ATR workflow that populates the model at scale.

The pilot comprises four components: the decision-unit data model with its controlled vocabularies; a structured dataset of 61 *Risposte*, with missing folio ranges modelled as first-class gap rows (register coverage ≈51%); the manual diplomatic transcriptions, which constitute the ground truth for ATR alignment and evaluation; and the rule-based extraction and quality-control procedures. Beyond the pilot lie the enlargement of the manually corrected ground truth, the training of a recognition model for the untranscribed portions, and the extension of the workflow to the full series.

The register presents three recurrent documentary markers: a date (including the intra-register *detto*, ditto), a marginal topic note (often present), and concluding signatories. These regularities ground rule-based segmentation and high-yield extraction. We define the *Risposta* as the primary decision unit and encode its recurring institutional variants (dissent, continuations) as explicit, typed relations, placing the model within the event- and relation-centric tradition of cultural-heritage data modelling (Doerr 2003).

The pilot also shows the limits of treating the unit as atomic. A single signed text may resolve several requests, quote its trigger (*supplica*: the petition), attach conditions to a grant, or speak with an individual rather than a collegial voice. The unit is nonetheless retained, on two grounds: the signed *Risposta* is the institutional act the Senate received, and in the domain expert's assessment it addresses a single matter in the large majority of cases. Composite units are therefore recorded as flagged exceptions within the unit; the planned refinement of the schema separates the documentary container from the deliberative acts it hosts, each with its own orientation and signatories.

The schema separates two layers. The minimal documentary layer captures source-near features required for segmentation, alignment, and automated population: register_id; unit_id folio_start/end; date_original and date_iso, marginal_note (with presence flag); signatories_raw/norm (with count/completeness); text_diplomatic; and transcription_status (manual_full / manual_partial / regest / not_transcribed). The relation vocabulary provides for dissent_to, continuation_of and repeated_supplication: a dissenting opinion forms a unit of its own, linked to the unit it contests, and a continuation points back to the block it extends. In register 142 the relation attested so far is repeated_supplication, linking the two entries of the Federici citizenship dossier. An analytical layer adds controlled vocabularies for research: policy_domain, decision_orientation, document_trigger, plus actors and geography for aggregation and network views. Values in this layer are hypotheses proposed by the analyst, recorded with an explicit uncertainty state and subject to expert validation. The two layers are released as two distinct datasets, joined on the unit identifier: the documentary dataset is stable and citable, while the analytical dataset is versioned separately, so that hypotheses can evolve without altering the record of the source.

Semi-automatic population and quality control rely on rule-based procedures in Python, a choice grounded in the register's language: stable and formulaic, it combines a Tuscan administrative base, a Venetian lexical component, and a Latin formulaic crust, a configuration that Venetian legal and administrative writing retains into the early eighteenth century (see Tomasin 2001, 2010). This stability makes high-yield extraction tractable with transparent rules. The pilot extracts dates, folio references, signatories, and monetary amounts and percentages by regular expression; institutional actors against a magistracy gazetteer; deontic formulas (*debba*: must, *non possa*: may not, *sotto pena*: under penalty, *stimiamo*: we deem) and a fiscal lexicon (*datij*: customs duties, *gravezze*: levies, *decima*: ten-per-cent tax, *tansa*: assessed levy, *ducati*: ducats) by controlled-vocabulary matching; and the cited norms and precedents, the *parti* (the resolutions) and statutes the Savi invoke, which the pilot is extracting as the basis for a future relation graph. TF-IDF and cosine similarity, preferred to neural embeddings for their transparency and their adequacy to a corpus of this size, support exploratory clustering and quality control, surfacing candidate groupings and outliers; they recover, for example, the two-entry Federici citizenship dossier as a most-similar pair. Reliability varies with the target: closed-class features (dates, monetary expressions, folio references) are high-yield targets, though orthographic variation and imperfect boundary detection still produce false positives, and all matches enter the dataset as candidates; interpretive features are structurally ambiguous, since bare modal verbs (*possa*: he may, *possano*: they may) can express normative permission or prohibition as well as epistemic possibility, and therefore enter the dataset as flagged hypotheses awaiting expert validation. The lexica, the controlled vocabularies, and the extracted hypotheses (each recorded with its rule identifier and its textual evidence) are available in the project's public repository; the extraction scripts are being consolidated for a subsequent release.[1]

Scaling Transcription: ATR, Interoperability, and Model Training

The second aim of the pilot is to assess the feasibility of a scalable transcription methodology, based on the services provided by H2IOSC, for the series in its entirety, thereby enabling large-scale analysis of institutional coherence, policy drift, and trade oversight in early modern Venice.

The Humanities and Cultural Heritage Italian Open Science Cloud (H2IOSC) federates the national consortia of the principal European research infrastructures for Cultural Heritage (E-RIHS-IT), Digital Humanities (DARIAH-IT), Computational Linguistics (CLARIN-IT), and Open Science (OPERAS-IT). H2IOSC is currently developing and testing a textual workflow built upon the services provided by each infrastructure. The pipeline originates from digital images delivered via the IIIF protocol (https://iiif.io), proceeds through Automated Text Recognition (ATR – a unified framework encompassing both HTR and OCR) and linguistic analysis via UDPipe, and concludes with the integration of bibliographic references through Zotero.

1 Project repository: https://github.com/vedph/risposte-savi; project site: https://vedph.github.io/risposte-savi/.

CLARIN-IT provides ATR services through the eScriptorium web platform (Kiessling et al. 2019; https://gitlab.com/scripta/escriptorium), originally developed by EPHE and now maintained with contributions from INRIA and other partners, as well as through new APIs integrated into the H2IOSC marketplace. To expand the range of ATR baseline models openly available to the scientific community, CLARIN-IT is implementing an interoperability strategy between Transkribus (Mühlberger et al. 2019; https://www.transkribus.org), accessed via a Team licence with a medium-scale credit allocation, and eScriptorium, connecting both platforms by routing outputs and inputs through the XML-PAGE interchange format. The two platforms play complementary roles. Transkribus currently offers higher out-of-the-box performance on generic documents through its supermodels (the Text Titan family, most recently Titan II), and is therefore a practical way to bootstrap a new transcription for manual correction; its models, however, are closed. eScriptorium, built on the open Kraken engine, allows project-specific models to be trained on the corrected material and released openly.

For this pilot the Text Titan I ter supermodel was employed; its output was imported into eScriptorium, where the automated transcription was aligned with the manual diplomatic transcription used as ground truth (Figure 1). The line-by-line segmentation model displayed an accuracy value of approximately 97%. For text recognition, a first project-specific model was created in eScriptorium from the aligned ground truth, which currently comprises twenty pages of diplomatic transcription: the model was trained on eighteen pages and evaluated on two held-out pages drawn from the same set, with no page used for both training and hyperparameter selection. The observed CER of ≈15% should be read as a provisional, indicative figure rather than a stable estimate of model performance, due to the small size of the sample: extending the manually corrected ground truth is the immediate priority for more reliable results.

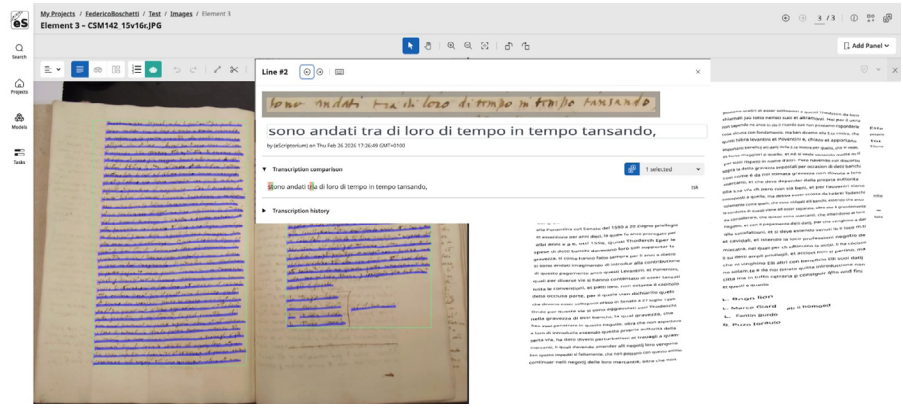

Figure 1: manual transcription merged with the automatic recognition (eScriptorium alignment view)

A new model will then be trained to recognise the portions of the manuscripts not yet manually transcribed. The model will be deposited in both the ILC4CLARIN repository and Zenodo, and its metadata will be made available through both the Virtual Language Observatory (VLO) of CLARIN and HTR-United.

Conclusions

The pilot delivers a documented model for segmenting and representing *Risposte* as decision units, a structured dataset comprising 61 records, built on manual diplomatic transcription as ground truth, a first evaluation of the ATR workflow (segmentation error ≈3%; text-recognition CER ≈15%), and the rule-based extraction and quality-control procedures released with the project repository. The next steps are the enlargement of the manually corrected ground truth, the training of a recognition model for the untranscribed portions, the act-level refinement of the schema, and the extension of the workflow to further registers. Together these operationalise the two directions of scalability: a stable decision-unit model (model scalability) and an interoperable transcription pipeline to populate it (quantity scalability).

Preliminary observations suggest recurring patterns across heterogeneous domains. The pilot generates a testable historical hypothesis: that decisions concerning citizenship, consular regulation, currency management, and maritime trade may share recurrent appeals to fiscal protection, harm containment, and procedural accountability. If confirmed on a larger and

independently validated corpus, these patterns would support an interpretation of the *Risposte* as components of a distributed governance architecture regulating economic flows. Extending the workflow to additional registers will make this hypothesis testable at scale, by quantifying segmentation accuracy, transcription quality and alignment effort, and precision and recall for high-yield extractions (dates, signatories, monetary expressions, deontic formulas).

Acknowledgements

The authors thank the Archivio di Stato di Venezia, CLARIN-IT and its national coordinator Monica Monachini for her support of the ATR workflow. The workflow relies on services developed within H2IOSC (Humanities and Cultural Heritage Italian Open Science Cloud), funded by the European Union – NextGenerationEU. Author contributions (CRediT): Emmanuela Carbé: conceptualization, methodology, data curation, software (project website and data-processing pipeline), writing – original draft, writing – review and editing. Federico Boschetti: methodology (ATR workflow), software, writing – original draft. Isabella Cecchini: investigation, resources (diplomatic transcription, photographic digitisation), validation, writing – original draft. The authors used generative AI to assist drafting, to prototype controlled vocabularies and extraction heuristics, and to develop the *Risposte* platform (Anthropic Claude Opus 4.8), with a final consistency check of the workflow (Anthropic Claude Fable 5); all outputs were reviewed and verified by the authors.

References

Borgherini-Scarabellin, M. (1925). *Il Magistrato dei Cinque Savi alla Mercanzia dalla istituzione alla caduta della Repubblica. Studio storico su documenti d'archivio*. Venezia: R. Deputazione.

Colavizza, G. (2017). A View on Venetian Apprenticeship through the Garzoni Database. In A. Bellavitis, M. Frank & V. Sapienza (eds.), *Garzoni. Apprendistato e formazione tra Venezia e l'Europa in età moderna*, 235-260. Mantova: Universitas Studiorum.

de Vivo, F. (2007). *Information and Communication in Venice: Rethinking Early Modern Politics*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199227068.001.0001

di Lenardo, I., Viaccoz, C., Guhennec, P., Musso, C. & Kaplan, F. (2025). The 1808 Napoleonic Land Registers of Venice: Spatial and Textual Dataset. *Journal of Open Humanities Data* 11: 67. DOI: 10.5334/johd.371

Doerr, M. (2003). The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine* 24(3): 75-92.

Flanders, J. & Jannidis, F. (eds.) (2019). *The Shape of Data in Digital Humanities: Modeling Texts and Text-based Resources*. Abingdon, Oxon-New York: Routledge.

Fusaro, M. (2015). *Political Economies of Empire in the Early Modern Mediterranean: The Decline of Venice and the Rise of England, 1450-1700*. Cambridge: Cambridge University Press.

Karch, T., Saydaliev, J., di Lenardo, I. & Kaplan, F. (2025). LLM Agents for Interactive Exploration of Historical Cadastre Data: Framework and Application to Venice. *Computational Humanities Research* 1: e11. DOI:10.1017/chr.2025.10014

Kiessling, B., Tissot, R., Stokes, P. & Stökl Ben Ezra, D. (2019). eScriptorium: An Open Source Platform for Historical Document Analysis. In *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, vol. 2, 19. IEEE. DOI: 10.1109/ICDARW.2019.10032

Mühlberger, G. et al. (2019). Transforming Scholarship in the Archives through Handwritten Text Recognition: Transkribus as a Case Study. *Journal of Documentation* 75(5): 954-976. DOI: 10.1108/JD-07-2018-0114.

Pezzolo, L. (2013). The Venetian Economy. In E. Dursteler (ed.), *A Companion to Venetian History, 1400-1797* (Brill's Companions to European History 4), 255-289. Leiden-Boston: Brill. DOI: 10.1163/9789004252523_007

Sella, D. (1968). Crisis and Transformation in Venetian Trade. In B. Pullan (ed.), *Crisis and Change in the Venetian Economy in the Sixteenth and Seventeenth Centuries*, 88-105. London: Methuen.

Sella, D. (1994). L'economia. In G. Cozzi & P. Prodi (eds.), *Storia di Venezia, VI: Dal Rinascimento al Barocco*, 651-711. Roma: Istituto della Enciclopedia Italiana.

Tomasin, L. (2001). *Il volgare e la legge. Storia linguistica del diritto veneziano (secoli XIII–XVIII)*. Padova: Esedra.

Tomasin, L. (2010). *Storia linguistica di Venezia*. Roma: Carocci.

Programme Committee:

Anna Maria Marras, PC Chair, University of Turin
Marina Buzzoni, Ca' Foscari University of Venice
Elisa Cugliana, University of Cologne
Swantje Dogunke, University of Jena
Rafał Górski, Jagiellonian University
Eetu Mäkela, University of Helsinki
Christopher Nunn, Heidelberg University
Petr Plecháč, Czech Academy of Sciences
Vassilis Routsis, University College London
Joëlle Weiss, Trier University

Local Organizing Committee:

Joanna Byszuk-Podsadniuk, Institute of Polish Language of the Polish Academy of Sciences
Maciej Eder, Institute of Polish Language of the Polish Academy of Sciences
Agata Kwaśnicka-Janowicz, Jagiellonian University
Wojciech Łukasik, Jagiellonian University
Adam Pawłowski, The University of Wrocław
Jan Rybicki, Jagiellonian University
Aleksandra Rykowska, Jagiellonian University
Michał Woźniak, Institute of Polish Language of the Polish Academy of Sciences