Acta Historica Tallinnensia 32/2 https://doi.org/10.3176/hist.2026.2.01

Medieval Studies after the Digital Turn: Methods, Infrastructures, and the Making of *Digital Livonia*

Marek Tamm

Tallinn University, School of Humanities, Institute of History, Archaeology and Art History, Narva mnt 25, 10120 Tallinn; marek.tamm@tlu.ee

Hembo Pagi

Tallinn University, School of Humanities, Institute of History, Archaeology and Art History, Narva mnt 25, 10120 Tallinn; hembo@tlu.ee

Abstract. This article examines the digital turn in medieval studies through a methodological survey and a reflective study of Digital Livonia, a regional research platform. It reviews developments in digital collections, editions, transcription, computational analysis, networks, GIS and artificial intelligence, then examines their application to dispersed sources, heterogeneous datasets and established research practices. It argues that integration requires selective formalisation: shared structures must enable comparison while preserving source distinctions, provisional identifications and the provenance of scholarly assertions. Digital Livonia demonstrates how data models, researcher workflows and institutional responsibility shape the possibilities and limits of digital historical enquiry.

Keywords: digital medieval studies, medieval Livonia, research infrastructures, digital editions, handwritten text recognition, computational text analysis, source criticism

© 2026 Authors. This is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International License CC BY 4.0 (http://creativecommons.org/licenses/by/4.0).

INTRODUCTION

Over the past two decades, the digital turn has been one of the most significant transformations in medieval studies. It has, however, not been the only major change to reshape the field. Global, material, environmental, postcolonial and gender-oriented approaches have altered its geographical horizons, its objects of study and many of its fundamental questions. Digital technologies have increasingly become part of the infrastructure through which such research is pursued. Their effects extend across almost every stage of the research process, from the discovery and reproduction of sources to their transcription, analysis, interpretation and publication. Even medievalists who would never describe themselves as digital humanists now work in an environment shaped by online catalogues, digitised manuscripts, searchable corpora, bibliographical databases and electronic communication. In this broad sense, digital resources have become part of the ordinary practice of medieval studies, including research that makes no use of computational analysis.[1]

The term 'digital turn' nevertheless risks suggesting a rupture where the historical development has been cumulative and uneven. Computers entered medieval research long before the World Wide Web, while printed catalogues, critical editions, archival visits and close reading continue alongside databases, digital facsimiles and machine learning. The digital also encompasses practices that involve very different degrees of formalisation and methodological intervention, from viewing a manuscript image or consulting an online repertory to training a model to identify scribal hands.[2] For analytical purposes, we distinguish three overlapping operations: digitisation makes existing sources and scholarly products available electronically; digital structuring represents them

1 For some general discussions, see D. J. Birnbaum, S. Bonde, M. Kestemont. The Digital Middle Ages: An Introduction. – Speculum, 2017, 92, S1, S1–S38. https://doi.org/10.1086/694236; R. Bleier, F. Fischer, T. Hiltmann, G. Viehhauser, G. Vogeler. Digitale Mediävistik und der deutschsprachige Raum. – Das Mittelalter, 2019, 24, 1, 1–12. https://doi.org/10.1515/mial-2019-0001; B. Burrichter, B. Gebert, C. Mackert, G. Viehhauser. Digitale Mediävistik. – Das Mittelalter, 2021, 26, 1, 101–117. https://doi.org/10.17885/heiup.mial.2021.1.24312; Medieval Manuscripts in the Digital Age. Ed. by B. Albritton, G. Henley, E. Treharne. Routledge, London, New York, 2021; Digital Medieval Studies: Practice and Preservation. Ed. by L. K. Morreale, S. Gilsdorf. Arc Humanities Press, Leeds, 2022. https://doi.org/10.1017/9781802700152; Digital Medieval Studies: Experimentation and Innovation. Ed. by S. Gilsdorf, L. K. Morreale. Arc Humanities Press, Leeds, 2024. https://doi.org/10.1515/9781802702330.

2 On identifying scribal hands, see, for instance, C. De Stefano, F. Fontanella, M. Maniaci, A. Scotto di Freca. Reliable Writer Identification in Medieval Manuscripts through Page Layout Features: The "Avila" Bible Case. – Engineering Applications of Artificial Intelligence, 2018, 72, 99–110. https://doi.org/10.1016/j.engappai.2018.03.023.

as searchable, linkable and reusable objects; computational analysis applies formal, statistical or algorithmic procedures to texts, images and historical data.[3] These distinctions concern what researchers do with evidence, rather than separate branches or successive stages of medieval studies. Search itself involves computation, and the production of a digital facsimile already requires choices about representation and description. They nevertheless provide a useful framework for examining how digital approaches have reorganised medieval research and how sources move from images and transcriptions to structured records and analytical corpora.

This introductory article develops that examination in two parts. The first part surveys selected areas of digital research in medieval studies, including digital collections and editions, manuscript transcription, computational textual analysis, relational and spatial methods, and artificial intelligence. The second examines the construction of Digital Livonia, a regional platform bringing together digitised sources, editions, thematic databases and research services. The survey establishes the methodological possibilities against which the choices and limitations of the platform can be assessed. Our argument develops from a practical problem encountered while building Digital Livonia: what should a regional research platform integrate, and what should remain specific to its constituent collections? Shared structures make evidence comparable, but they can also obscure differences between documentary attestations, uncertain identifications and modern reconstructions. Conversely, preserving every collection in its existing form limits the connections that researchers can establish between them. We examine how this tension shaped the platform's data model, technical architecture and working procedures. These choices form part of what we call a hybrid practice of medieval research, in which digital procedures are developed and assessed through sustained engagement with philology, palaeography, codicology and historical source criticism. Collaboration between historians, technical specialists and collection custodians helps to maintain an identifiable relationship between historical assertions and the evidence on which they depend.

The other contributions to this special issue examine these questions through different bodies of Livonian evidence. Nicolas Bourgeois assesses statistical and machine-learning approaches to the *Chronicle*

3 On the distinction between digitisation, digital methods and digital history, see C. Brennan. Digital Humanities, Digital Methods, Digital History, and Digital Outputs. – History Compass, 2018, 16, 10, e12492. https://doi.org/10.1111/hic3.12492.

of Henry of Livonia and the *Livonian Rhymed Chronicle*, considering how computational analysis can support verifiable interpretation and collaboratively curated metadata. Rūta Bruževica follows the transformation of craft ordinances from Riga and Tallinn into a digital corpus, exploring how textual analysis and visualisation depend on transcription, classification and the interpretation of normative sources. Turning to material evidence, Anu Mänd demonstrates how reflectance transformation imaging and digital photomontage can improve the reading of worn grave slabs and reconnect dispersed fragments, including through comparison with historical drawings. Juhan Kreem places digital imaging within the longer history of reproducing medieval Livonian seals, showing how successive techniques have shaped their study and why the proliferation of images needs to be accompanied by adequate scholarly metadata. The final two articles examine the manuscript and archival contexts on which digital research also depends. Tiina Kala and Kaspar Kolk investigate the textual composition, binding and liturgical fragments of the *Ex quo* vocabulary manuscript Cm 16 in the Tallinn City Archives, distinguishing its connections with the Dominican friary from the unresolved identity and affiliation of its scribe. Kristjan Toomaspoeg reconstructs the history of research on Livonica in the Vatican Archives, with particular attention to Aleksander Kurtna, and identifies the need to bring dispersed editions, transcriptions and research papers into a common database. Together, these studies show how the possibilities of digital medieval research are shaped by the transmission of texts and objects, the history of their documentation and the scholarly decisions through which they become available for comparison.

FROM HUMANITIES COMPUTING TO DIGITAL MEDIEVAL STUDIES

The history of digital medieval studies is commonly narrated through the Italian Jesuit Roberto Busa and his *Index Thomisticus*. From his doctoral research in the mid-1940s, Busa sought a means of indexing every occurrence of every word in the works of Thomas Aquinas and related authors; during a visit to the United States in 1949, he secured the support of IBM. Yet, as Steven E. Jones has shown, this familiar account functions as a founding myth and should not be mistaken for a simple

point of origin.[4] Busa did not step directly into modern computing. His early work depended on punch-card machines developed for business data processing and is more accurately described as humanities data processing than as digital humanities in the present sense. The project grew out of a particular post-war conjunction of Jesuit scholarly networks, IBM's technical support, engineers such as H. Paul Tasman, and extensive, often unacknowledged, data labour, much of it performed by women.[5] Its importance lies both in the unprecedented scale of the resulting eleven-million-word concordance and in this socio-technical form of scholarship: text was decomposed into units, encoded according to explicit rules and reassembled, while the available machinery helped to determine what could be asked and produced.

Busa's project was exceptional in scale, but it should not obscure the multiple beginnings of computer-assisted medieval research. Edgar Lejeune's recent research on European medievalists in the 1960–1990 period shows that the field emerged in several rather different settings. At the Centre de traitement électronique des documents (CETEDOC) in Louvain, Paul Tombeur and his collaborators developed automated concordances of medieval Latin texts. From 1966, David Herlihy and Christiane Klapisch-Zuber organised a transatlantic project to turn the Florentine *catasto* of 1427 into a database for demographic and social analysis. In the 1970s, Jean-Philippe Genet and his colleagues created the ALINE program and the MEDITEXT corpus for lexicometric analysis. These enterprises represented their sources in different ways – as full texts, selected variables, documentary descriptors, concordances or programs – but all were collective, bringing historians together with demographers, linguists, computer scientists, punch-card operators and institutional computing centres. Lejeune's central contribution is to locate the transformation in what he calls the 'practices of sources': their selection and criticism, the constitution of corpora and the operations by which evidence was synthesised. The Florentine *catasto* project, initially described as an 'edition for the computer', is exemplary. Its data were not simply present in the manuscript waiting to be extracted; they were produced through decisions about variables,

4 S. E. Jones. Roberto Busa, S. J., and the Emergence of Humanities Computing: The Priest and the Punched Cards. Routledge, New York, 2016, 1–29, 107–173.

5 See J. Nyhan, A. Flinn. Computation and the Humanities: Towards an Oral History of Digital Humanities. Springer, Cham, 2016, esp. 59–62, 79–80, 90. https://doi.org/10.1007/978-3-319-20170-2; M. Terras, J. Nyhan. Father Busa's Female Punch Card Operatives. – Debates in the Digital Humanities 2016. Ed. by M. K. Gold, L. F. Klein. University of Minnesota Press, Minneapolis, 2016, 60–65. https://doi.org/10.5749/j.ctt1cn6thb.9.

categories, omissions, correction and uncertainty. Early computer use thus demanded a new degree of formalisation and made particularly visible that data are constructed forms of evidence. It also prompted collective organisation beyond individual projects. A meeting in Rome in 1975 led to the foundation in 1979 of the review *Le Médiéviste et l'ordinateur*, whose debates about workflows, open or closed research instruments, common standards, reuse and technological obsolescence remain strikingly current.[6]

By the 1990s, projects such as the Electronic Beowulf and the Piers Plowman Electronic Archive were testing what an electronic scholarly environment could offer that print could not.[7] Manuscript images, transcriptions, descriptions and apparatus could now be brought into direct relation, while the Text Encoding Initiative (TEI), founded in 1987, supplied a shared framework for representing textual structures. Personal computing, optical media and especially the web moved such work beyond institutional computing centres and specialised laboratories.[8] Yet many of the questions raised by the earlier generation (standards, reuse, collaboration and technological obsolescence) remained unresolved.[9]

Since the 2010s, the growth of digitised collections, interoperable image services, handwritten text recognition, computer vision and machine learning has expanded the possibilities for connecting research operations. Earlier projects also sought to integrate databases, corpora and editions, while many contemporary resources remain isolated. What has changed is the availability of shared standards and services through which materials can move between digitisation, description,

6 E. Lejeune. Médiévistes et ordinateurs: Organisations collectives, pratiques des sources et conséquences historiographiques (1966–1990). Thèse de doctorat d'histoire et de philosophie des sciences, Université de Paris, 2021. https://theses.hal.science/tel-03682773v2. See also E. Lejeune. 'Open' or 'Close' Research Instruments? Conflicting Rationales in the Organization of Early Digital Medieval History in Europe (1960–1990). – Digital Humanities Quarterly, 2024, 18, 2. https://doi.org/10.63744/f7ddtgjkbjyw; E. Lejeune. From Sources to Data: Medievalists, Demographers, and Computer Scientists Producing a Database from a Quattrocento Census (1964–1978). – History of Humanities, 2025, 10, 2, 397–425. https://doi.org/10.1086/737038.

7 A. Prescott. The Electronic Beowulf and Digital Restoration. – Literary and Linguistic Computing, 1997, 12, 3, 185–196. https://doi.org/10.1093/llc/12.3.185; The Piers Plowman Electronic Archive, Vol. 1: Corpus Christi College, Oxford, MS 201 (F). Ed. by R. Adams, H. N. Duggan, E. Eliason, R. Hanna III, J. Price-Wilkin, T. Turville-Petre. University of Michigan Press for the Society for Early English and Norse Electronic Texts, Ann Arbor, 2000.

8 J. Unsworth. Medievalists as Early Adopters of Information Technology. – Digital Medievalist, 2012, 7. https://doi.org/10.16995/dm.34.

9 P. Robinson. Current Issues in Making Digital Editions of Medieval Texts – or, Do Electronic Scholarly Editions Have a Future? – Digital Medievalist, 2005, 1. https://doi.org/10.16995/dm.8; S. P. McCormick. A Guide to Digital Medieval Studies in North America. – Perspectives médiévales, 2016, 37. https://doi.org/10.4000/peme.9655; P. Sahle. What Is a Scholarly Digital Edition? – Digital Scholarly Editing: Theories and Practices. Ed. by M. J. Driscoll, E. Pierazzo. Open Book Publishers, Cambridge, 2016, 19–39. https://doi.org/10.11647/OBP.0095.02.

transcription, annotation, analysis and publication.[10] Donald J. Waters has described this as the emergence of an 'end-to-end' digital infrastructure for the humanities.[11] Such an infrastructure remains incomplete and unevenly distributed, but the concept draws attention to the connections between research operations and the work required to sustain them.[12] The developments briefly outlined here did not replace one another, and their boundaries remain porous. The history of the field is better understood as a cumulative and uneven attempt to integrate changing research instruments with the older practices of source criticism and historical interpretation.[13]

DIGITAL COLLECTIONS AND RESEARCH INFRASTRUCTURES

For working medievalists, the most immediate effect of digitisation has been access. Manuscripts, charters, seals, inscriptions and other sources held in different countries can now be compared within the same working session. This has changed the geography of research as well as its speed. Materials once accessible mainly to scholars with substantial travel budgets and institutional connections can increasingly be consulted from elsewhere.[14] The International Image Interoperability Framework (IIIF) has pushed this development beyond the provision of images. Because compatible repositories can expose images through a common protocol, a manuscript held in one collection can be viewed, compared and annotated alongside material held elsewhere. Codices, fragments and documentary series dispersed by the history of collecting can therefore be provisionally reunited in a common research environment.[15]

10 C. L. Borgman, A. Scharnhorst, M. S. Golshan. Digital Data Archives as Knowledge Infrastructures: Mediating Data Sharing and Reuse. – Journal of the Association for Information Science and Technology, 2019, 70, 8, 888–904. https://doi.org/10.1002/asi.24172.

11 D. J. Waters. The Emerging Digital Infrastructure for Research in the Humanities. – International Journal on Digital Libraries, 2023, 24, 87–102, here 88–90. https://doi.org/10.1007/s00799-022-00332-3.

12 U. Pawlicka-Deger. Infrastructuring Digital Humanities: On Relational Infrastructure and Global Reconfiguration of the Field. – Digital Scholarship in the Humanities, 2022, 37, 2, 534–550. https://doi.org/10.1093/llc/fqab086; C. L. Borgman, A. Scharnhorst, M. S. Golshan. Digital Data Archives as Knowledge Infrastructures.

13 P. Svensson. The Landscape of Digital Humanities. – Digital Humanities Quarterly, 2010, 4, 1. https://doi.org/10.63744/5dhf3ymfs2gg.

14 A. Cappellotto. From Codex to Apps: The Medieval Manuscript in the Age of Its Digital Reproduction. – Umanistica Digitale, 2020, 4, 9, 1–18. https://doi.org/10.6092/issn.2532-8816/11459; T. Burrows. The History and Provenance of Manuscripts in the Collection of Sir Thomas Phillipps: New Approaches to Digital Representation. – Speculum, 2017, 92, S1, S39–S64. https://doi.org/10.1086/693438.

15 R. Sanderson, B. Albritton, R. Schwemmer, H. Van de Sompel. SharedCanvas:

Yet access remains uneven. Digitisation is expensive and selective, and well-funded or visually attractive collections are more likely to appear online than poorly catalogued materials. The image itself is also a transformation of the object: scale, weight, texture, sequence and traces of handling are difficult to reproduce on screen.[16] Advanced imaging can also reveal features unavailable in conventional photographs. Multispectral imaging, reflectance transformation imaging and 3D modelling can disclose erased writing, surface relief and traces of pigment that may be difficult or impossible to perceive through ordinary inspection.[17] For seals, inscriptions and other material sources, lighting direction and image processing can make a worn letter or shallow impression more legible, while also shaping its appearance. Researchers therefore need documentation of the conditions of capture and processing and, where possible, access to alternative views of the same surface. These allow another scholar to assess a proposed reading and consider how the imaging procedure may have enhanced, obscured or distorted the feature under examination. Catalogue metadata require the same scrutiny as images. Inherited classifications and uncertain attributions can acquire a misleading precision when presented as database fields; documenting their origin and revision allows users to distinguish institutional descriptions from subsequent scholarly interpretations.[18]

Our experience with Digital Livonia has made clear how much this critical work depends on infrastructure. Preserving images, documenting their production and keeping research data available for scrutiny and reuse require repositories, standards, software maintenance and sustained institutional commitments. FAIR provides a useful vocabulary – findable, accessible, interoperable and reusable – but achieving these qualities requires continuing scholarly work.[19]

A Collaborative Model for Digital Facsimiles. – International Journal on Digital Libraries, 2012, 13, 1, 3–16. https://doi.org/10.1007/s00799-012-0098-8; L. F. Davis. Framing the Fragment, from Parchment to Pixel. – Digital Medievalist, 2024, 17, 1, 1–23. https://doi.org/10.16995/dm.10343.

16 H. Bamford, E. Francomano. On Digital-Medieval Manuscript Culture: A Tentative Manifesto. – Digital Philology, 2018, 7, 1, 29–45. https://doi.org/10.1353/dph.2018.0002; L. F. Davis. Life On – and Off – the Continuum. – Digital Medieval Studies: Practice and Preservation, 99–111. https://doi.org/10.1017/9781802700152.006.

17 B. Endres. Digitizing Medieval Manuscripts: The St. Chad Gospels, Materiality, Recoveries, and Representation in 2D and 3D. Arc Humanities Press, Leeds, 2019.

18 P. Andrist. Toward a New Generation of Databases and Database Applications for Describing Ancient Manuscripts. – Digital Scholarship in the Humanities, 2021, 36, Supplement 2, ii9–ii16. https://doi.org/10.1093/llc/fqab022.

19 M. D. Wilkinson et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. – Scientific Data, 2016, 3, 160018. https://doi.org/10.1038/sdata.2016.18.

DIGITAL EDITIONS AND THE MODELLING OF TEXTUAL TRADITIONS

Scholarly editing is among the areas in which medieval sources have posed particularly productive challenges to digital methods. Medieval works often survive in witnesses that differ in wording, order, language, layout and material context. Printed editions have developed sophisticated ways of representing this plurality, including synoptic texts and extensive apparatuses, but the relations between these elements remain constrained by the organisation of the page and volume. Digital editions permit different arrangements of images, diplomatic and normalised transcriptions, translations, variants and commentary. Their value depends on which relations they allow readers to inspect and how clearly they disclose the editorial decisions behind them.[20]

A digital edition constitutes an explicit model of the textual tradition. The editor must decide what counts as a textual unit, how abbreviations and corrections are represented, whether uncertain readings are retained, how witnesses are aligned and which relations between text and material object are made visible. Encoding languages such as TEI render many of these decisions explicit and machine-readable. They allow the same data to generate different views: a close transcription beside the facsimile, a normalised reading text, a synoptic comparison of witnesses or a corpus for linguistic analysis. The scholarly argument therefore lies not only in the introductory essay and reconstructed text, but also in the structure of the data.[21] For historians working with charters, inventories or account books, another question concerns the relationship between an edited text and the information extracted from it. Georg Vogeler's concept of the 'assertive edition' addresses this relation by linking textual evidence to structured statements about people, objects, events and their connections.[22]

Digital editions are particularly well suited to what has been called the *mouvance* or variance of medieval textuality.[23] Instead of treating differences between witnesses mainly as corruption on the

20 E. Pierazzo. Digital Scholarly Editing: Theories, Models and Methods. Ashgate, Farnham, 2015, 5–24. https://doi.org/10.4324/9781315573564.

21 P. Sahle. What Is a Scholarly Digital Edition?

22 G. Vogeler. The 'Assertive Edition': On the Consequences of Digital Methods in Scholarly Editing for Historians. – International Journal of Digital Humanities, 2019, 1, 309–322. https://doi.org/10.1007/s42803-019-00025-5.

23 P. Zumthor. Essai de poétique médiévale. Seuil, Paris, 1972, 164, 507; B. Cerquiglini. Éloge de la variante: Histoire critique de la philologie. Seuil, Paris, 1989, 54.

way to an original, they can represent textual change as an object of study in its own right. Gabriel Viehhauser's survey of digital editing in medieval German studies shows both the potential and the limits of this development. Projects such as the digital edition of *Der arme Heinrich* combine manuscript images, several levels of transcription, translation and information about the codex, thereby restoring a degree of visibility to the plurality of the tradition. At the same time, the number of fully developed digital medieval editions remains smaller than repeated predictions of a digital revolution might lead one to expect.[24] These limitations arise in large part from the institutional, financial and technical demands of digital scholarly editing. Digital editions require sustained collaboration between editors, libraries and research software specialists; they are costly to develop and demanding to maintain. Interfaces age faster than printed books, links decay and project-specific software may become unusable. Moreover, academic evaluation still tends to recognise the finished monograph or critical edition more readily than the creation of data, code or an evolving scholarly resource. The digital edition has consequently revealed an institutional problem as much as a technical one: scholarship increasingly depends on collective and continuing labour, while systems of authorship, funding and credit remain oriented towards individual and final products.[25]

MAKING MANUSCRIPTS MACHINE-READABLE

Digitisation has created an abundance of manuscript images, but images are not searchable textual corpora. For most medieval sources, transcription remains the principal bottleneck. Optical character recognition (OCR) developed for modern print performs poorly on handwriting, abbreviations and irregular layouts. Handwritten text recognition (HTR) addresses the problem through supervised machine learning: a model is trained on images paired with corrected transcriptions, learns recurring visual patterns and predicts characters on pages that were not included in the training material. As Peter Stokes

24 G. Viehhauser. Zum gegenwärtigen Stand der digitalen Editorik in der Altgermanistik: Einige Bemerkungen. – Computer im Musenhain: Von träumenden Büchern und der Aura des Digitalen. Hrsg. von S. Bartsch, L. Borek, P. Hegel. Technische Universität Darmstadt, Darmstadt, 2023, 15–25. https://doi.org/10.26083/tuprints-00027905.

25 L. K. Morreale, S. Gilsdorf. Introduction: The Medievalist, Digital Edition. – Digital Medieval Studies: Practice and Preservation, 1–10. https://doi.org/10.1017/9781802700152.001; D. J. Waters. Emerging Digital Infrastructure, 96–99.

and Benjamin Kiessling emphasise, the performance and range of such systems depend fundamentally on the data available to train them.[26] Progress has consequently been fastest for modern, widely used languages and scripts, whereas many historical writing systems lack corpora of sufficient size and consistency. The selection and preparation of training data already determine which materials become amenable to automated study and which remain comparatively inaccessible.[27]

The practical importance of HTR is considerable. Output need not be publication-ready to support full-text search, the identification of names and formulae, or the selection of passages for correction. A relatively small set of manually transcribed pages can make a much larger collection searchable. General models for Latin and vernacular scripts are also reducing the quantity of project-specific training material required, although performance varies sharply between hands, genres, layouts and states of preservation. The increasing circulation of reusable models and training datasets may reduce duplicated labour, particularly for comparatively rare scripts. Their reuse nevertheless requires common standards, or at least explicitly documented conventions, for transcription and data exchange.[28]

HTR has returned an old philological question to the centre of attention: what should a transcription record? Expanding abbreviations, modernising punctuation or normalising spelling may aid reading but erase features relevant to scribal practice. Conversely, recording every glyph can produce data that are difficult to reuse. Estelle Guéville and David Joseph Wrisley argue that transcription schemes must follow the research question. Their work on Paris Bibles preserves abbreviation and orthographic variation because these features help distinguish scribes and regional habits. The 'ground truth' used to train a model is therefore already an interpretation: transcription determines which features of the manuscript become available for subsequent computational analysis and

26 P. A. Stokes, B. Kiessling. Sharing Data for Handwritten Text Recognition (HTR). – The Companion to Digital Humanities in Practice. Ed. by C. Crompton, L. Estill, R. J. Lane, R. Siemens. Routledge, London, New York, 2025, 95–104. https://doi.org/10.4324/9781003327677-8.

27 G. Mühlberger et al. Transforming Scholarship in the Archives through Handwritten Text Recognition: Transkribus as a Case Study. – Journal of Documentation, 2019, 75, 5, 954–976. https://doi.org/10.1108/JD-07-2018-0114; A. Pinche, P. A. Stokes. Historical Documents and Automatic Text Recognition: Introduction. – Journal of Data Mining and Digital Humanities, 2024. https://doi.org/10.46298/jdmdh.13247.

28 A. Pinche, T. Clérice. Reconnaissance automatique d'écriture et sources historiques: Limites et nouvelles perspectives. – Studi Francesi, 2025, 206, 357–366. https://doi.org/10.4000/15gi8; D. J. Wrisley, E. Guéville. Medieval Manuscripts and the Computational Humanities: Big Data, Scribes, and the "Paris Bible". Arc Humanities Press, Leeds, 2026, 43–82. https://doi.org/10.17302/MFSK1586.

which disappear through normalisation.[29] This problem has also been examined by Elena Pierazzo and Peter Stokes, who argue that manuscript transcription should preserve the relationship between text and codex rather than extract a purely linguistic sequence from its material setting. Pages, columns, lines, additions, erasures and spatial arrangements are not incidental containers of textual content; depending on the question being asked, they may form part of the evidence that the transcription is intended to represent.[30] Transcription is consequently a model of the manuscript, created through decisions about which dimensions of the object are retained and which are omitted.

Ariane Pinche's research on early French vernacular legendaries illustrates how improvements in HTR can extend the corpus available for philological investigation. Her account moves from an initial experiment on a single manuscript towards the use of recognition models on further witnesses. The resulting transcriptions provide material for comparing the composition of hagiographical collections and investigating the possible circulation of smaller groups of saints' lives before their incorporation into larger compilations.[31] A comparable expansion of the available evidence is occurring in visual and material analysis, where digitisation increasingly means producing new forms of evidence from reflected light, surface structure and comparison across images. Computer vision can identify page regions, measure proportions of text and decoration, compare letterforms or search for recurring motifs. The eCodicology project analysed around 170,000 images from manuscripts associated with the Benedictine abbey of St Matthias in Trier, extracting layout features in order to identify patterns across the collection.[32] DigiPal and related projects led by Stokes have similarly used digital images and structured descriptions of letterforms to support comparison, scribal attribution and the study of scripts. Such approaches render

29 E. Guéville, D. J. Wrisley. Transcribing Medieval Manuscripts for Machine Learning. – Journal of Data Mining and Digital Humanities, 2024. https://doi.org/10.46298/jdmdh.9805.

30 E. Pierazzo, P. A. Stokes. Putting the Text Back into Context: A Codicological Approach to Manuscript Transcription. – Kodikologie und Paläographie im digitalen Zeitalter 2. Codicology and Palaeography in the Digital Age 2. Ed. by F. Fischer, C. Fritze, G. Vogeler. Books on Demand, Norderstedt, 2010, 397–430; P. A. Stokes. Holistically Modelling the Medieval Book: Towards a Digital Contribution. – Anglia, 2021, 139, 1, 6–31. https://doi.org/10.1515/ang-2021-0002.

31 A. Pinche. Comprendre la composition des premiers légendiers en langue vernaculaire à l'aide des méthodes numériques: Acquisition automatique du texte (HTR) et analyses stylométriques. – Bulletin du centre d'études médiévales d'Auxerre, 2022, Collection CBMA. https://doi.org/10.4000/cem.19452.

32 H. Busch, S. Chandna. eCodicology: The Computer and the Mediaeval Library. – Kodikologie und Paläographie im digitalen Zeitalter 4. Hrsg. von H. Busch, F. Fischer, P. Sahle. Books on Demand, Norderstedt, 2017, 3–23; B. Endres. Digitizing Medieval Manuscripts.

selected palaeographical features measurable and comparable, while their identification and the interpretation of resulting patterns continue to depend on palaeographical judgement.[33]

The possibilities opened by HTR and related forms of image analysis are accompanied by problems of uneven reliability. In HTR, errors are not random: models may perform well on regular main text but fail on marginalia, unusual abbreviations, changes of hand or damaged passages – precisely the features that may matter most. Accuracy percentages also conceal the difference between common and historically significant errors.[34] To make these limitations assessable, published transcriptions should document their provenance, identify the recognition model and transcription conventions used, and distinguish uncorrected output from text reviewed by researchers, with a record of subsequent corrections. Where recognition accuracy is reported, the evaluation sample and its relation to the wider collection should also be specified. The dependence of recognition quality on the material condition of the source is especially visible in the Rescapé project on the fire-damaged Turin manuscript BNU L.II.14. Severe scorching initially prevented automatic segmentation because the contrast between ink and parchment was insufficient, while tears, cuts and warped text lines produced further recognition problems. The project therefore combined specialised imaging with complementary tools for region and line detection and adapted its transcription conventions to represent damaged and illegible text. The case locates HTR within a longer chain extending from conservation and imaging through segmentation and transcription to interoperable data.[35]

33 P. A. Stokes. Scribal Attribution across Multiple Scripts: A Digitally-Aided Approach. – Speculum, 2017, 92, S1, S65–S85. https://doi.org/10.1086/693968; S. Brookes, P. A. Stokes, M. Watson, D. Marques de Matos. The DigiPal Project for European Scripts and Decorations. – Writing Europe, 500–1450: Texts and Contexts. Ed. by A. Conti, O. Da Rold, P. Shaw. D. S. Brewer, Cambridge, 2015, 25–58.

34 M. J. Hill, S. Hengchen. Quantifying the Impact of Dirty OCR on Historical Text Analysis: Eighteenth Century Collections Online as a Case Study. – Digital Scholarship in the Humanities, 2019, 34, 4, 825–843. https://doi.org/10.1093/llc/fqz024; T. Clérice, M. Vlachou-Efstathiou, A. Chagué. CREMMA Medii Aevi: Literary Manuscript Text Recognition in Latin. – Journal of Open Humanities Data, 2023, 9, 4. https://doi.org/10.5334/johd.97.

35 G. Kagan. Numériser des manuscrits brûlés. Techniques et méthodes à l'aide de l'intelligence artificielle: nouvelles perspectives. – Studi Francesi, 2025, 206, 335–340. https://doi.org/10.4000/15gi5; P. O'Connor. Handwritten Text Recognition Technology and MS Turin, BNU, L.II.14 (T). The "Rescapé" Case Study. – Studi Francesi, 2025, 206, 367–376. https://doi.org/10.4000/15gi9.

READING TEXTS AT SCALE

Once texts are machine-readable, they can be compared across corpora beyond the limits of ordinary linear reading. Computational textual analysis includes word frequencies, stylometry, authorship attribution, text-reuse detection, topic modelling and semantic methods. Their common premise is that patterns too dispersed or repetitive for unaided reading may become visible when textual features are counted and compared systematically.[36] An earlier experiment in Pinche's research on vernacular legendaries, conducted jointly with Jean-Baptiste Camps and Thibault Clérice, combined HTR and stylometry to test Paul Meyer's hypotheses about the grouping of saints' lives in Paris, BnF, fr. 412. Working with texts copied by a single hand helped to limit scribal variation, although recognition errors and differences between the texts still complicated the analysis. The resulting groupings provided evidence for assessing the proposed compilations and possible shared authorship, with their interpretation requiring comparison with the manuscript tradition.[37]

Stylometry is one form of this work. It measures recurrent features such as frequent words, character sequences, vocabulary or syntax to compare texts and has been applied to anonymous authorship, scribal intervention, genre and dialect. Medieval transmission complicates the inference from statistical similarity to authorship because the surviving text may preserve interventions by compilers, translators, scribes and modern editors.[38] A comparison of letters by Hildegard of Bingen, Guibert of Gembloux and Bernard of Clairvaux suggested that Guibert substantially reworked two visionary texts transmitted under Hildegard's name.[39] Quantitative spelling analysis of *The Canterbury Tales* has confirmed palaeographical identifications of manuscripts copied by the

36 See, for instance, F. Stella. Digital Philology and Quantitative Criticism of Medieval Literature: Unconventional Approaches to Medieval Latin Literature II. Brepols, Turnhout, 2021. https://doi.org/10.1484/M.USML-EB.5.119334; J. De Gussem. Computational Stylistics and Medieval Texts. – Routledge Resources Online: Medieval Studies. Ed. by H. Klemettilä, S. Niskanen, J. Willoughby. Routledge, 2022, 1–12. https://doi.org/10.4324/9780415791182-RMEO391-1.

37 J.-B. Camps, T. Clérice, A. Pinche. Noisy Medieval Data, from Digitized Manuscript to Stylometric Analysis: Evaluating Paul Meyer's Hagiographic Hypothesis. – Digital Scholarship in the Humanities, 2021, 36, Supplement 2, ii49–ii71. https://doi.org/10.1093/llc/fqab033.

38 D. J. Wrisley, E. Guéville. Medieval Manuscripts and the Computational Humanities, 89–116; S. Dows-Miller. Medieval Authorship and Statistical Analysis: Reconsidering Jehan de Saint-Quentin's 'œuvre' in the Digital Age. ARC Humanities Press, Leeds, 2026, 17–28.

39 M. Kestemont, S. Moens, J. Deploige. Collaborative Authorship in the Twelfth Century: A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux. – Digital Scholarship in the Humanities, 2015, 30, 2, 199–224. https://doi.org/10.1093/llc/fqt063.

same scribes,[40] while word embeddings applied to the DEEDS corpus have traced changes in administrative vocabulary across the Norman Conquest. The questions differ, but each depends on converting textual variation into features that can be compared systematically.[41] Corpus analysis and close reading operate most effectively in succession. The first locates anomalies and tests representativeness; the second evaluates what a pattern may mean. Often the process is recursive.

Such analysis is complicated by the particular character of medieval corpora. They are frequently small, multilingual, editorially heterogeneous and unrepresentative; spelling, dates and attributions are unstable. Latin and canonical authors remain easier to model than many vernacular or documentary traditions. Differences in transcription, editorial intervention and corpus composition must therefore be examined before statistical patterns are attributed to medieval authors or scribes.[42]

RELATIONS AND PLACES: PROSOPOGRAPHY, NETWORKS AND GIS

Charters, registers, letters, chronicles and administrative records repeatedly connect persons, institutions, offices, events and places. Digital prosopography makes it possible to identify such entities across sources and investigate them collectively, while network analysis and geographic information systems (GIS) provide ways of examining the relations and spatial patterns that emerge from these data.

These relations, however, are not simply given by the sources. A line in a network graph is already an interpretation. Co-witnessing a charter, kinship, correspondence, monastic membership and denunciation in an inquisitorial record arise from different documentary practices and cannot be treated as interchangeable evidence of 'connection'. A graph is not a picture of medieval society, but a model constructed from selected

40 P. M. W. Robinson, T. Tresoldi. Analyzing Spelling Patterns in the Manuscripts of the *Tales of Canterbury*. – Digital Scholarship in the Humanities, 2026, 41, 1, 347–363. https://doi.org/10.1093/llc/fqaf123.

41 Y. Liu, G. Tilahun, X. Gao, Q. Wen, M. Gervers. A Comparative Study of Static and Contextual Embeddings for Analyzing Semantic Changes in Medieval Latin Charters. – Proceedings of the First Workshop on Language Models for Low-Resource Languages. Association for Computational Linguistics, Abu Dhabi, 2025, 182–192. https://aclanthology.org/2025.loreslm-1.14/.

42 On the inseparability of corpus construction and interpretation, see K. Bode. A World of Fiction: Digital Collections and the Future of Literary History. University of Michigan Press, Ann Arbor, 2018, 17–44. https://doi.org/10.3998/mpub.8784777; D. J. Wrisley, E. Guéville. Medieval Manuscripts and the Computational Humanities.

sources and decisions about nodes and ties. Measures such as centrality, clustering or brokerage acquire historical meaning only in relation to the evidence from which those relations were derived.[43] This dependence on documentary form makes source criticism part of the network model itself. An apparent difference between the network positions of men and women, for example, may reflect the composition of surviving testimonies, the interests of investigators or the circumstances under which information was recorded.[44] Such factors can sometimes be encoded and tested rather than left as external qualifications. Zbíral, Sikk and Shaw have proposed precisely this kind of integration of digital methods and source criticism in their work on inquisitorial records.[45] Courtney Luckhardt's study of women's networks in Carolingian Septimania shows a complementary possibility: relational analysis can recover actors and connections obscured by narratives organised around office, dynasty or male political agency.[46] Studies of inquisitorial material likewise reconstruct dissident communities while asking how interrogation, denunciation and documentary procedure produced the network preserved in the source.[47]

Mapping Medieval Chester reconstructed the urban landscape around 1500 and set it beside English, Latin and Welsh textual representations of the city. The comparison made it possible to examine how the same urban space was perceived and narrated within a multilingual frontier community.[48] Justin Colson confronted a different problem in late medieval and early modern London, where deeds and wills located properties through relational descriptions such as 'next to', 'opposite'

43 N. Ruffini-Ronzani, S. de Valeriola, E. Cuvelier. Dealing with the Heterogeneity of Interpersonal Relationships in the Middle Ages: A Multi-Layer Network Approach. – Digital Medievalist, 2022, 15, 1. https://doi.org/10.16995/dm.8070; S. de Valeriola. Can Historians Trust Centrality? Historical Network Analysis and Centrality Metrics Robustness. – Journal of Historical Network Research, 2021, 6, 1, 85–125. https://doi.org/10.25517/jhnr.v6i1.105; K. Sikk, R. Välimäki, D. Zbíral. How Inquisitive Was Medieval Inquisition? A Network-Analytical Approach to Information Flow in the Trials for Brandenburg-Pomeranian Waldensians (Late 14th c.). – Digital Scholarship in the Humanities, 2026, 41, 1, 396–411. https://doi.org/10.1093/llc/fqaf138.

44 D. Salihović, J. L. Estévez. Gender Bias in Medieval Inquisitions and Its Place in Shaping Knowledge about the Heterodox. – Social Science History, 2026, 50, 1, 171–201. https://doi.org/10.1017/ssh.2025.10099.

45 D. Zbíral, K. Sikk, R. L. J. Shaw. Bridging Digital History Methods and Source Criticism: A Research Agenda for the Study of Inquisition Records. – Acta Historica Tallinnensia, 2025, 31, 1, 46–72. https://doi.org/10.3176/hist.2025.1.02.

46 C. Luckhardt. Mapping Women's Relational Networks in Carolingian Septimania, 795–850. – Early Medieval Europe, 2025, 33, 3, 412–435. https://doi.org/10.1111/emed.12779.

47 D. Zbíral, K. Riccardo, T. Hampejs, Z. Brys. Gender, Kinship, and Other Social Predictors of Incrimination in the Inquisition Register of Bologna (1291–1310): Results from an Exponential Random Graph Model. – PLOS ONE, 2025, 20, 2, e0315467. https://doi.org/10.1371/journal.pone.0315467.

48 K. D. Lilley. Urban Mappings: Visualizing Late Medieval Chester in Cartographic and Textual Form. – Mapping the Medieval City: Space, Place and Identity in Chester, c. 1200–1600. Ed. by C. A. M. Clarke. University of Wales Press, Cardiff, 2011, 19–41.

and 'between'. Translating these descriptions into GIS-compatible data produced a relative urban topography from evidence that had never been organised according to modern cartographic coordinates.[49] These projects illustrate a recurring problem in historical GIS: spatial analysis requires the formalisation of historical descriptions without imposing a precision foreign to the sources. Geocoded data can reveal density, proximity, accessibility and change over time, but coordinates are themselves scholarly propositions.[50] David Joseph Wrisley's mapping of place names in medieval French literature extends the same principle from reconstructed physical space to textual geography, allowing the geographical horizons of genres, circulation and literary production to be examined comparatively.[51]

Neither relational nor spatial formalisation removes uncertainty. A medieval name may identify several people or places, dates may remain ranges, and a proposed location may be only approximate. Data models need to preserve these alternatives and degrees of uncertainty. Historical gazetteers, linked data and knowledge graphs can connect variant names and competing identifications across projects, provided that interoperability does not erase local distinctions or conceal the evidential basis and degree of certainty attached to each assertion.[52]

ARTIFICIAL INTELLIGENCE, RETRIEVAL AND THE LIMITS OF GENERAL-PURPOSE MODELS

Artificial intelligence (AI) belongs within the longer history of computational medieval studies. Machine learning already underlies HTR, image classification, named-entity recognition and linguistic analysis. Generative AI has changed the terms of engagement by allowing

49 J. Colson. Reinterpreting Space: Mapping People and Relationships in Late Medieval and Early Modern English Cities Using GIS. – Urban History, 2020, 47, 3, 384–400. https://doi.org/10.1017/S0963926820000164.

50 See R. L. J. Shaw, K. Sikk, D. Zbíral. Toponymic Surnames and the Spatiality of Heresy Prosecutions: Peter Seila's Register of Sentences from the Quercy Region (Languedoc), 1241–1242. – Humanities and Social Sciences Communications, 2024, 11, 195, 1–14. https://doi.org/10.1057/s41599-024-02689-z.

51 D. J. Wrisley. Locating Medieval French, or Why We Collect and Visualize the Geographic Information of Texts. – Speculum, 2017, 92, S1, S145–S169. https://doi.org/10.1086/694300.

52 R. Simon, L. Isaksen, E. Barker, P. de Soto Cañamares. The Pleiades Gazetteer and the Pelagios Project. – Placing Names: Enriching and Integrating Gazetteers. Ed. by M. L. Berman, R. Mostern, H. Southall. Indiana University Press, Bloomington, 2016, 97–109. https://doi.org/10.2307/j.ctt2005zq7.12; A. Mertel, D. Zbíral, Z. Stachoň, H. Hořínková. Historical Geocoding Assistant. – SoftwareX, 2021, 14, 100682. https://doi.org/10.1016/j.softx.2021.100682.

scholars to interact with systems that produce fluent text, translations, summaries and code. Large language models can support exploratory searching, language work, information extraction and the creation of research tools, especially for scholars without advanced programming skills.[53]

For medieval studies, however, fluency is a poor proxy for historical reliability. Relevant languages and textual traditions are unevenly represented in training data, manuscript variation and editorial mediation are typically obscured in general-purpose systems, and plausible answers may contain fabricated references or silently modernised readings. Translation can smooth ambiguity and reproduce assumptions inherited from modern editions. Commercial systems create an additional problem because their provenance and reproducibility are difficult to establish.[54] Specialised language technologies can offer greater control over these procedures. Frameworks designed for pre-modern languages may document their models, preprocessing choices and linguistic assumptions and permit evaluation against defined corpora. Specialisation alone, however, guarantees neither transparency nor reproducibility. These depend on access to the relevant components, stable versions, adequate documentation and an evaluation appropriate to the task. The same criteria should guide the assessment of general-purpose models, whose usefulness for a particular operation must be established through testing rather than inferred from the fluency of their output.[55]

Retrieval over a curated corpus offers a more readily assessable experimental use of generative AI. A system can search a delimited collection of transcriptions, catalogue records or editions and return identifiable passages alongside a generated synthesis. This makes some errors easier to diagnose, although a cited passage may still fail to support the claim attached to it, and relevant evidence may be missed

53 Y. Fox. Using Generative AI in Historical Practice. Cambridge University Press, Cambridge, 2026. https://doi.org/10.1017/9781009740067.

54 On the fundamental limits of large language models as sources of knowledge, see E. M. Bender, T. Gebru, A. McMillan-Major, S. Shmitchell. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? – Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. ACM, New York, 2021, 610–623. https://doi.org/10.1145/3442188.3445922; W. H. Walters, E. I. Wilder. Fabrication and Errors in the Bibliographic Citations Generated by ChatGPT. – Scientific Reports, 2023, 13, 14045. https://doi.org/10.1038/s41598-023-41032-5; É. Ollion, R. Shen, A. Macanovic et al. The Dangers of Using Proprietary LLMs for Research. – Nature Machine Intelligence, 2024, 6, 4–5. https://doi.org/10.1038/s42256-023-00783-6.

55 K. P. Johnson, P. J. Burns, J. Stewart, T. Cook, C. Besnier, W. J. B. Mattingly. The Classical Language Toolkit: An NLP Framework for Pre-Modern Languages. – Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations. Association for Computational Linguistics, Online, 2021, 20–29. https://doi.org/10.18653/v1/2021.acl-demo.3.

during retrieval. For Digital Livonia we have so far kept general-purpose conversational AI outside the public interface. Our preferred direction is to investigate retrieval over curated collections, evaluating the selection of evidence separately from the accuracy of any generated answer and retaining links to records, passages and source images.

A similar emphasis on assessability appears in pedagogical uses of generative AI. Sophie Whittle's work on Chaucer asks students to compare AI-generated translations with manuscripts, editions and human-produced alternatives, identifying omissions, biases and anachronisms. Here, error itself becomes an object of criticism and a means of teaching editorial judgement. The model's output functions as material for comparison rather than as an authoritative interpretation.[56] A more research-oriented version of the same principle can be found in structured information extraction. When language models identify and classify evidence according to a closed annotation scheme, their performance can be evaluated systematically. A reproducible workflow records the model and its version, minimises stochastic variation and compares the results with a manually created gold-standard sample. Precision and recall then provide more meaningful measures than the apparent fluency of individual answers. Experiments with medieval inquisition records have used this procedure to distinguish testimony from procedural language, identify persons and groups, classify their grammatical and narrative roles, and assign reported actions to predefined semantic categories. Such evaluation leaves the model's internal opacity intact while placing empirical bounds on the reliability of its output.[57]

Across these examples, the location of interpretative work shifts. Decisions once concentrated in editions and historical arguments now also enter corpus construction, transcription schemes, data models and model evaluation. Computational methods alter the scale and organisation of comparison; their historical value depends on the evidential relations they make visible. For us, this shift requires criticism at three stages: the source itself, its transformation into data and the tools applied to those data. Zbíral, Sikk and Shaw's notion of 'source criticism 2.0' goes further by incorporating documentary production,

56 S. Whittle. Chaucer in the Age of AI: Critical and Collaborative Translation in Digital Medieval Studies. – Digital Culture & Education, 2026, 16, 3, 67–90. See also M. Schöffel, M. Wiedner, E. Garces Arias, P. Ruppert, C. Heumann, M. Aßenmacher. Modern Models, Medieval Texts: A POS Tagging Study of Old Occitan. – Proceedings of the 5th International Conference on Natural Language Processing for Digital Humanities. Ed. by M. Hämäläinen, E. Öhman, Y. Bizzoni, S. Miyagawa, K. Alnajjar. Association for Computational Linguistics, Albuquerque, 2025, 334–349. https://doi.org/10.18653/v1/2025.nlp4dh-1.30.

57 D. Zbíral, K. Sikk, R. L. J. Shaw. Bridging Digital History Methods and Source Criticism.

transmission, perspective and survival into computational analysis itself.[58] The Digital Livonia case that follows makes this redistribution of scholarly work visible in practice.

BUILDING A REGIONAL KNOWLEDGE INFRASTRUCTURE: THE DIGITAL LIVONIA PLATFORM

No single research platform can encompass all six domains outlined above. Digital Livonia concentrates on collections, editions, transcription, entity management and spatial access, while computational textual analysis and network analysis are principally downstream uses of exported data; general-purpose generative AI has not been incorporated into the public environment. Its contribution therefore lies less in embedding every analytical method than in creating documented and reusable points of passage between sources, structured data and external research tools.

Digital Livonia was established at Tallinn University in 2021 to create a common research environment for the history of medieval Livonia. Funded by the Estonian Research Council until 2025, the project brought together historians, art historians, archaeologists, manuscript specialists and researchers in digital technologies.[59] Its principal public outcome is a web platform available in both Estonian and English, combining digitised sources, digital source editions, research databases, practical applications, images of material objects and a growing set of services for searching, filtering, mapping and reusing them.[60] As participants in the conception and construction of Digital Livonia, we examine decisions, revisions and abandoned assumptions that are rarely visible in the finished interface. Our account draws on this experience rather than an independent evaluation of the platform's usability or research impact.

58 D. Zbíral, K. Sikk, R. L. J. Shaw. Bridging Digital History Methods and Source Criticism; R. L. J. Shaw, K. Suba, T. Hampejs, D. Zbíral. Syntactic-Semantic Capture of Historical Texts as a Platform for Source-Critical Analysis: Telling the Story of a Premodern Heresy Trial with Computer-Assisted Semantic Text Modelling (CASTEMO). – Digital Scholarship in the Humanities, 2026, 41, 2, 1001–1020. https://doi.org/10.1093/llc/fqag016.

59 The project (PRG1276) was led by Marek Tamm, and the core team included Rūta Bruževica, Inna Jürjo, Villu Kadakas, Tiina Kala, Kaspar Kolk, Juhan Kreem, Ivar Leimus, Anu Mänd, Hembo Pagi, and Karl Peeter Valk.

60 Digital Livonia: For a Digitally Enhanced Study of Medieval Livonia (c. 1200–1550), https://dl.tlu.ee/en/ (accessed 20 September 2026). Unless otherwise stated, all descriptions of the platform's current contents and functionality refer to the version consulted on that date.

The platform formed one component of a broader research programme organised around social and prosopographical studies, spatial and network analysis, intellectual and manuscript history, and economic history. The thematic databases and digital editions were developed within these research strands rather than produced according to a single platform-wide schema. This organisation helps to explain both the scholarly richness of the resulting collections and the differences between their structures, vocabularies and levels of completion.

A REGION CONSTITUTED BY DISPERSED EVIDENCE

Medieval Livonia presents a particularly revealing case for such an undertaking because the historical region no longer corresponds to a modern political, linguistic or institutional unit. Its territory is divided chiefly between Estonia and Latvia, while relevant sources are held in archives, libraries and museums across the Baltic region, Germany, the Vatican, Poland, Sweden, Denmark, Russia and elsewhere. Their dispersal reflects the political and ecclesiastical organisation of medieval Livonia, its connections with the Hanseatic world, the papacy and the religious orders, as well as the subsequent movement of archives, libraries and private collections. The study of the region has consequently depended on crossing modern state borders and combining source traditions organised according to different national historiographies and archival practices.

The surviving evidence is heterogeneous even within a single institution. The Tallinn City Archives, for example, preserves municipal registers, account books, correspondence, property records, guild statutes, wills and manuscript codices, whereas the National Archives of Estonia holds charters, seal impressions and various materials concerning landed property and ecclesiastical institutions. Museums preserve seals, coins, inscriptions, liturgical objects and works of art; libraries hold manuscript fragments and printed editions; archaeological records add buildings, urban plots, castles and excavated objects. Much of this material has been catalogued and/or digitised, but digitisation has usually proceeded according to the remit and technical system of the holding institution. A charter may therefore appear in one repository, its seal in another database, its printed edition in a nineteenth-century volume held by a

digital library, and the persons or places mentioned in it in a separately compiled scholarly repertory.

The fragmentation of sources affects both access and the kinds of questions researchers can readily pursue. Institutional catalogues generally reflect the boundaries of particular collections and source types, while historical research often requires connections across them. A study of a Tallinn merchant, for example, may draw on municipal registers, probate inventories, property transactions, correspondence, guild membership records, artworks and material held in foreign archives. Research into a cleric's intellectual life may similarly connect prosopographical information with manuscripts, university attendance, ecclesiastical offices and the circulation of particular texts. Where these sources can be searched only through separate institutional or disciplinary systems, such connections are harder to identify and their reconstruction demands considerable time and scholarly effort.

Digital Livonia was conceived in response to this problem. Its object was not to create a comprehensive digital copy of the 'medieval Livonian archive', which would have been technically, legally and intellectually unrealistic. The more practicable aim was to establish a common point of access where selected digital collections, newly prepared source editions and research data could be described according to compatible principles and, where possible, connected through shared historical entities. In this respect the project treated medieval Livonia as a field of evidence rather than a fixed territorial container. The coherence of the platform would derive from the relationships between people, places, institutions, events, documents and objects, even when the underlying resources remained in different repositories. Such an approach accords with a broader tendency in recent Livonian historiography to replace an image of the region as an isolated periphery with an analysis of the networks through which people, texts, objects and practices circulated.[61]

The platform's regional remit nevertheless exceeds its present coverage. Many of its collections concern Tallinn or the territory of present-day Estonia, and the public interface is available in Estonian and English. Although several datasets also include evidence from Latvia, this

61 See, for instance, M. Tamm. Deprovincialising Livonia: St. Catherine's Convent of Tallinn in Thirteenth-Century Dominican Networks. – Livland – eine Region am Ende der Welt? Forschungen zum Verhältnis zwischen Zentrum und Peripherie im späten Mittelalter / Livonia – a Region at the End of the World? Studies on the Relations between Centre and Periphery in the Later Middle Ages. Ed. by A. Selart, M. Thumser. (Quellen und Studien zur baltischen Geschichte, 27.) Böhlau, Köln, Weimar, Wien, 2017, 61–90; Making Livonia: Actors and Networks in the Medieval and Early Modern Baltic Sea Region. Ed. by A. Mänd, M. Tamm. Routledge, New York, London, 2020. https://doi.org/10.4324/9780429296000.

distribution reflects the location of the project team, its research interests and its institutional partnerships. Digital Livonia should therefore be understood as a growing contribution to the study of the region, with substantial geographical imbalances that need to remain visible to users. Extending coverage also requires decisions about collaboration, editorial responsibility and multilingual access.

FROM PROPOSAL TO PLATFORM

The first plans for Digital Livonia were ambitious. The platform was expected to aggregate digitised resources from several memory institutions, host new prosopographical and thematic databases, publish searchable source editions, support maps and network analysis, enrich imported metadata and provide a semantic search across different source types. The envisaged data model was to be compatible with the CIDOC Conceptual Reference Model, while Dublin Core and controlled vocabularies would provide a basis for metadata exchange. Several interfaces were imagined: a research environment for scholars, a more accessible public version and, eventually, educational or gamified applications. These aims were intellectually coherent, yet they assumed a degree of technical integration and standardisation that could not be achieved simply by selecting an existing content-management system.

An early formulation described Digital Livonia as a data project rather than a software project. The distinction expressed an important concern. Technologies age quickly, while historical data should remain usable after a particular interface or application has disappeared. A small research group cannot make the continued maintenance of a bespoke information system its principal scholarly task. The proposed solution was therefore to keep the data as independent as possible from the public website, to rely on open standards and established tools, and to store digitised materials in institutional repositories with stronger preservation infrastructures. These principles have remained central. The expectation that an adequate platform could be assembled largely from ready-made components, however, proved optimistic.

The difficulty became apparent when the project moved from general entities such as ‘person’, ‘document’ or ‘object’ to the structure of actual sources. A person mentioned in a municipal account book is not documented in the same manner as a knight brother whose career has been reconstructed from charters and office lists. A book may survive

as a physical codex, be known from a probate inventory, appear as a title reconstructed by a modern scholar, or combine all three conditions. Prices and wages introduce units, currencies, commodities, dates and conversion problems that bear little resemblance to the description of manuscripts or castles. A generic schema broad enough to accommodate every collection risked becoming too abstract for meaningful research, while a separate structure for each dataset would reproduce the fragmentation that the platform was intended to overcome.

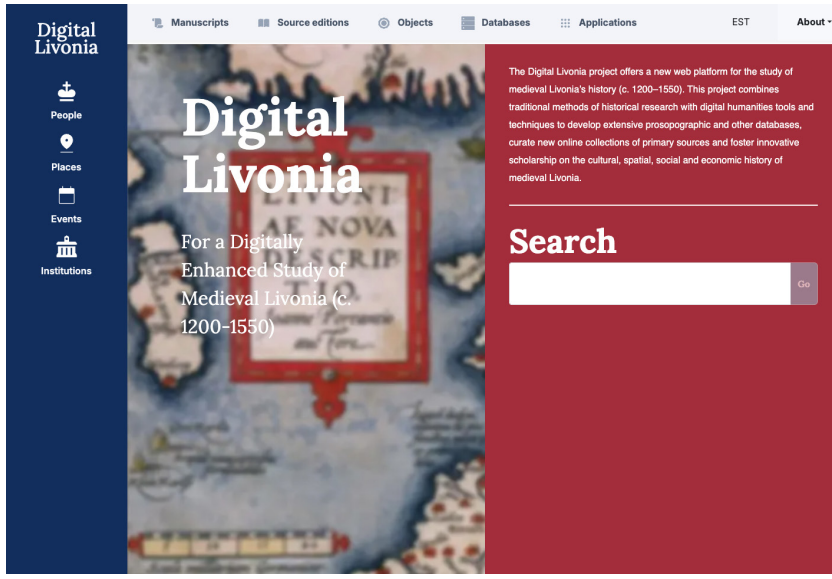
During development, we moved from planning comprehensive semantic integration to implementing connections incrementally. Digitised images, transcriptions, thematic databases, maps and entity indexes could become usable before every relation between them had been resolved. The present system relies principally on relational structures, controlled vocabularies and selected links between records. CIDOC CRM remains a conceptual reference for defining relations across collections, but a complete mapping was beyond the available resources. This was a decision about the extent and timing of implementation: relational storage can support source-specific assertions and subsequently expose selected information through CRM mappings or linked-data services.

This change reflects the need to accommodate developments in research as well as a move towards a simpler model. As scholarship advances, new sources may require unforeseen categories, identifications may be revised, and previously accepted relations may become uncertain. Fields designed to address one research question may also prove inadequate for another. A platform for historical research therefore needs enough structure to support comparison, together with the flexibility to revise its data model. The initial Digital Livonia model was understood as a working hypothesis, open to revision through further research. The resulting architecture aims to make existing collections usable and to support connections between them, without assuming that a single, comprehensive model can adequately represent the historical world to which they relate.

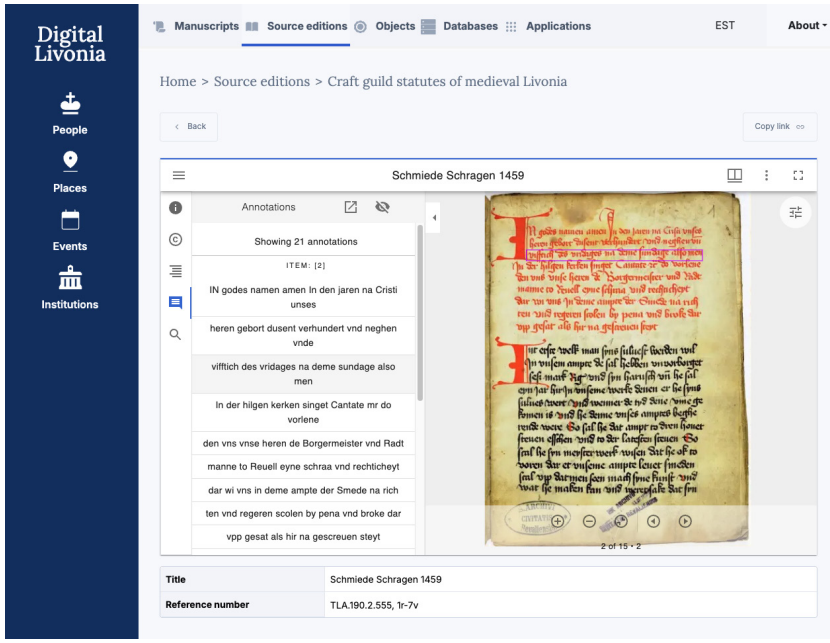
SOURCES, EDITIONS AND RESEARCH DATA

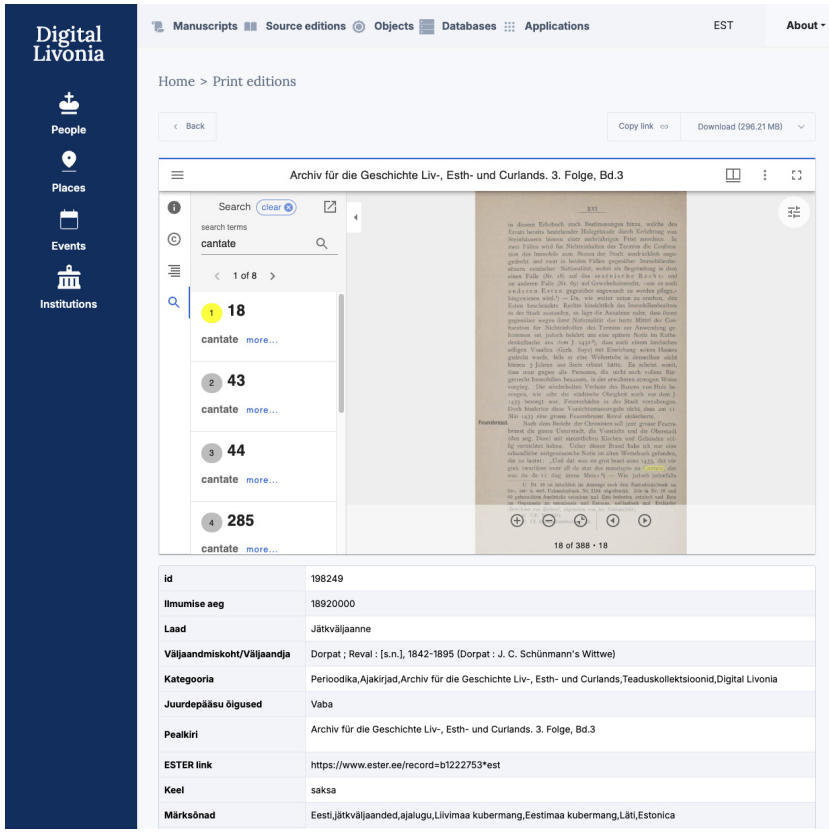
The public platform reflects this layered development. At the time of writing, its contents are organised into five principal sections: manuscripts, source editions, objects, databases and applications. The manuscript section contains seven collections, including the membership books of the Tallinn Brotherhood of the Black Heads, parchment charters and letters, manuscript fragments gathered under *Fragmenta membranea Estoniae*, Tallinn account books, registers of burghers, property and mortgage registers, and manuscript codices. The source-edition section comprises the probate inventories of medieval Tallinn, the statutes of craft guilds from Tallinn and Riga, the civic ordinances known as *burspraken* and digitised printed editions of Livonian sources. A separate object collection presents medieval Livonian seals.[62]

These categories describe different forms of digital representation. Some collections lead primarily to digitised images accompanied by catalogue information. Others combine images with manually prepared transcriptions and segmentation, while some editions use HTR in Transkribus to produce an initial transcription that is subsequently checked and corrected by researchers. The collection of printed source editions draws heavily on the digitisation infrastructure of the Tallinn



62 Digital Livonia, Manuscripts, https://dl.tlu.ee/en/parchments/; Source editions, https://dl.tlu.ee/en/source-editions/; Objects, https://dl.tlu.ee/en/objects/ (accessed 6 September 2026).


Digital Livonia
People
Places
Events
Institutions
Manuscripts
Source editions
Objects
Databases
Applications
EST
About
Home > Source editions > Craft guild statutes of medieval Livonia
Back
Copy link
Schmiede Schragen 1459
Annotations
Showing 21 annotations
ITEM: [2]
IN godes namen amen In den jaren na Cristi unses
heren gebort dusent verhundert vnd neghen vnde
vifftich des vridages na deme sundage also men
In der hilgen kerken singet Cantate mr do vorlene
den vns vnse heren de Borgermeister vnd Radt
manne to Reuell eyne schraa vnd rechticheyt
dar wi vns in deme ampte der Smede na rich
ten vnd regeren scolen by pena vnd broke dar
vpp gesat als hir na gescreuen steyt
2 of 15 • 2
Title
Schmiede Schragen 1459
Reference number
TLA.190.2.555, 1r-7v


Digital Livonia
People
Places
Events
Institutions
Manuscripts
Source editions
Objects
Databases
Applications
EST
About
Home > Print editions
Back
Copy link
Download (296.21 MB)
Archiv für die Geschichte Liv-, Esth- und Curlands. 3. Folge, Bd.3
Search
clear
search terms
cantate
1 of 8
1 18 cantate more...
2 43 cantate more...
3 44 cantate more...
4 285 cantate more...
18 of 388 • 18
id
198249
Ilmumise aeg
18920000
Laad
Jätkväljaanne
Väljaandmiskoht/Väljaandja
Dorpat ; Reval : [s.n.], 1842-1895 (Dorpat : J. C. Schünmann's Wittwe)
Kategooria
Perioodika,Ajakirjad,Archiv für die Geschichte Liv-, Esth- und Curlands,Teaduskollektsioonid,Digital Livonia
Juurdepääsu õigused
Vaba
Pealkiri
Archiv für die Geschichte Liv-, Esth- und Curlands. 3. Folge, Bd.3
ESTER link
https://www.ester.ee/record=b1222753*est
Keel
saksa
Märksõnad
Eesti,jätkväljaanded,ajalugu,Liivimaa kubermang,Eestimaa kubermang,Läti,Estonica

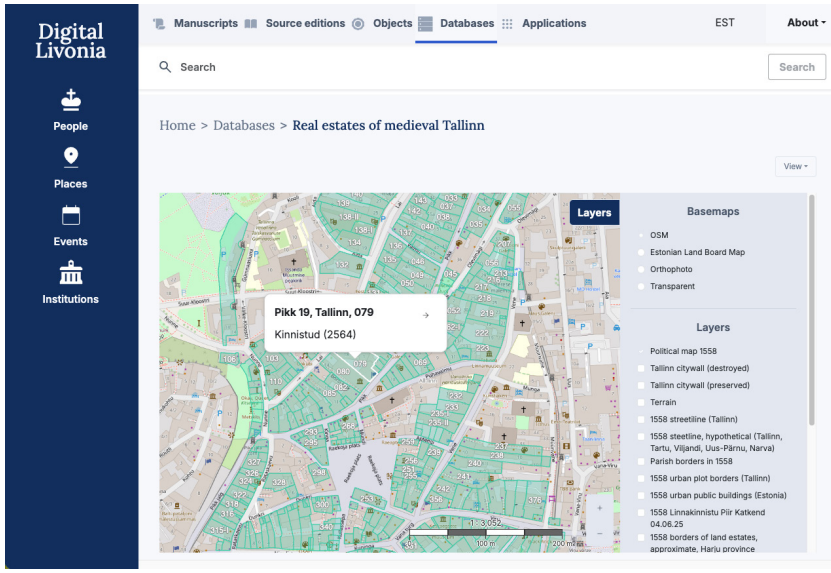


Figure 1. Selected interfaces of the Digital Livonia platform

University Academic Library and makes large nineteenth- and early twentieth-century publications available as searchable digital objects. The probate inventories, *burspraken* and guild statutes approach a digital edition more closely, since the images, transcribed text and structural divisions of the documents have been prepared together. The platform thus contains resources situated at several points between digital facsimile, searchable corpus, source edition and structured dataset. This diversity is appropriate to the sources, although it also means that the degree of searchability and analytical preparation varies from one collection to another.

Thirteen thematic databases are currently available. They concern real estate in medieval Tallinn, inhabitants of small medieval towns in present-day Estonia, inscriptions on Livonian artworks, prices, Tallinn scribes, books, castles, wages, the knight brethren of the Livonian branch of the Teutonic Order, Livonian clergy, manuscript codices, higher clergy and medieval Estonian manors.[63] Their subjects correspond to distinct fields of medieval research, yet together they illustrate the range of entities and relations that a regional platform must accommodate. The database of knight brethren reconstructs careers through offices

63 Digital Livonia, Databases, https://dl.tlu.ee/en/databases/ (accessed 20 September 2026). Since the individual databases continue to be supplemented, their descriptions should be understood as statements of their current coverage rather than claims to completeness.

and places, while the database of higher clergy records episcopal and capitular positions. The database of inhabitants of medieval towns contains prosopographical information on residents of towns in present-day Estonia, excluding Tallinn, covering eight chartered towns as well as Keila, Kuressaare and Lihula. The book database reconstructs Livonian book culture by combining surviving codices and early printed editions with documentary references to volumes that have since been lost, whereas the manuscript-codices database provides detailed codicological and textual descriptions of selected extant manuscripts. The inscription database relates verbal texts to material objects; the price and wage databases formalise amounts, currencies, commodities and labour; and the databases of castles, manors, and urban real estate link historical evidence to spatial locations and allow the data to be explored on an interactive map. The current scope and operational status of the databases are summarised in Table 1.

Table 1. The thematic databases currently available in Digital Livonia as of 20 September 2026

DATABASE	PRINCIPAL ENTITIES AND INFORMATION RECORDED	PUBLIC SCOPE AND PRESENT COVERAGE
Real estate of medieval Tallinn	Urban plots, owners and property transactions	Medieval Tallinn; coverage continues to expand
Inscriptions on medieval Livonian artworks	Inscribed texts, objects, languages and locations	Selected artworks from medieval Livonia; expanding corpus
Prices in medieval Livonia	Commodities, monetary amounts, currencies and dates	Coverage determined by the surviving source series
Wages in medieval Livonia	Forms of labour, payments, currencies and dates	Coverage determined primarily by surviving account books
The scribes of medieval Tallinn	Scribes, hands, documents and municipal offices	Documentary culture of medieval Tallinn
Knight brethren of the Livonian branch of the Teutonic Order	Persons, offices, careers and places	Livonian branch of the Teutonic Order

Table 1. Continued

DATABASE	PRINCIPAL ENTITIES AND INFORMATION RECORDED	PUBLIC SCOPE AND PRESENT COVERAGE
Medieval Livonian higher clergy	Bishops, canons, ecclesiastical offices and careers	Higher ecclesiastical officeholders in medieval Livonia; expanding corpus
Medieval Livonian clergy	Bishops, canons, priests, monks, friars, ecclesiastical offices and careers	Broader coverage of secular and regular clergy in medieval Livonia
Books in medieval Livonia	Surviving codices and early printed editions, together with documentary references to lost volumes	Medieval Livonian book culture; expanding corpus
Manuscript codices in medieval Livonia	Codicological and textual descriptions of surviving manuscripts	Selected Livonian manuscript collections; expanding corpus
Castles in medieval Livonia	Fortified sites, variant names, locations and historical information	Medieval Livonia; expanding corpus
Medieval Estonian manors	Sites, names, ownership and chronology	Territory of present-day Estonia; expanding corpus
Inhabitants of medieval Estonian small towns	Persons, origins, ownership, places	Towns in the territory of present-day Estonia; expanding corpus

Note: All databases listed were publicly operational on 20 September 2026; several remain under active expansion.

The platform currently includes two specialised applications that illustrate the formalisation required by digital historical work. A calculator for medieval Livonian currencies takes account of the fact that monetary values cannot be converted through a single stable ratio, since accounting systems and the silver content and relationships of units changed over time; it therefore distinguishes chronological periods and allows users to convert between the mark, ferding, schilling, öre, artig and pfennig according to historically defined equivalences.[64] A second

64 Digital Livonia, Medieval Livonian currency calculator, https://dl.tlu.ee/en/calculator/ (accessed 20 September 2026).

application, the weight calculator, translates between the principal weight units used in Livonian towns, using the Riga weight system as a common reference point while accounting for local variation in their actual values. It thus makes historically variable systems of measurement directly comparable across Riga, Tallinn, Tartu/Narva, Pärnu and Viljandi, turning a potentially cumbersome set of conversion rules into a reusable research tool.[65]

Several collections offer tabular interfaces through which users can search records, select columns, sort results, change views and download data. These functions are modest when compared with advanced analytical environments, but they mark an important transition from consultation to reuse. A fixed catalogue page presents the institution's description of an object; a configurable and downloadable table allows researchers to reorganise a corpus according to their own questions and carry it into statistical, network or geographical software. Digital Livonia does not perform every possible analysis within the platform. It seeks instead to provide sufficiently structured data from which further analysis can begin. The same principle informs our general indexes of people and places.[66] The people interface brings together records from several databases in a searchable and downloadable table, while the places section presents place records on a map organised by basemaps, layers and data sources. Our aim is to develop these interfaces into an entity-centred platform in which a person or place can provide a route between documents, objects and datasets. Cross-platform search for events and institutions remains under development, and links between the individual collections are being established progressively.

Digital Livonia currently supports parts of a source-centred workflow. The alignment of transcriptions with images allows readers to examine textual readings against the digitised document, while downloadable tables make selected research datasets available for further analysis. Movement between an individual source and related entries in other collections remains less consistent: some connections are supplied through links and indexes, whereas others require manual searching and historical identification. Following a charter through its image, description, named persons and associated objects remains a task shared between the platform and the researcher.

65 Digital Livonia, Medieval Livonian weight unit calculator, https://dl.tlu.ee/en/weight-calculator (accessed 20 September 2026).

66 Digital Livonia, People, https://dl.tlu.ee/en/databases/people/; Places, https://dl.tlu.ee/en/places/ (accessed 4 August 2026).

ARCHITECTURE AND INTEROPERABILITY

The technical architecture was designed around a separation between data management and public presentation. The principal data layer uses Directus, a headless content-management system placed above a relational database. A headless system stores and manages content without prescribing the form in which it must be displayed. The public frontend is a separate web application that requests data through an application programming interface. Directus generates REST and GraphQL API interfaces from the underlying database structure, including its relational fields.[67] The same data can consequently be used by the current website, a future interface, a map service or an external research application without being duplicated for each presentation. Figure 2 summarises the resulting division of responsibilities. Master files and repository metadata remain with memory institutions and ETERA; image conversion and delivery, transcription, structured data management and public presentation are handled by separate but connected services.

Separating these components allows the public interface to be redesigned without necessarily rebuilding the research databases, and permits other applications to use the same records. It also creates dependencies that require continuing maintenance: interfaces, permissions and software versions must remain compatible. Modularity therefore facilitates replacement and migration but does not itself

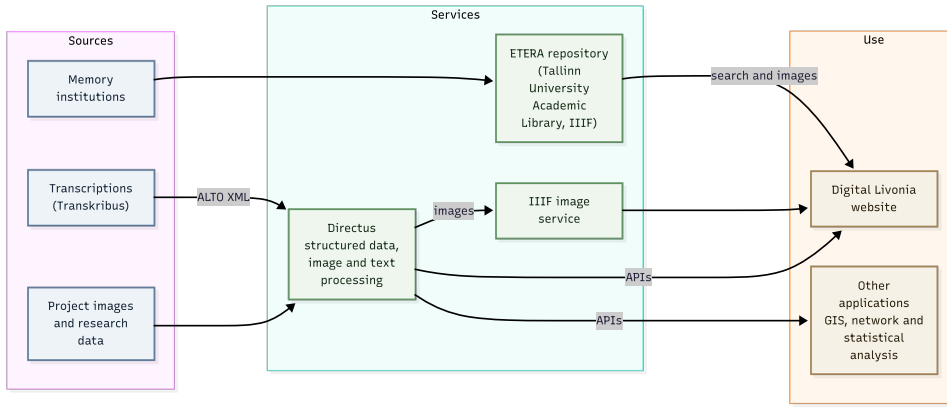


Figure 2. The modular architecture and principal data flows of Digital Livonia

67 Directus. Headless CMS. – Directus Documentation, https://directus.com/docs/getting-started/overview; Directus. API Reference. – Directus Documentation, https://directus.com/docs/api. REST – Representational State Transfer, https://en.wikipedia.org/wiki/REST; GraphQL – an open-source query language for APIs and a server-side runtime, https://graphql.org/ (accessed 20 September 2026).

preserve the data. Long-term reuse also requires durable exports, documented schemas and vocabularies, identifiable releases and an institution responsible for retaining them. These requirements apply especially to the project's original research datasets, whose preservation is distinct from the institutional custody of source images.

The data model had to represent sources that exist in several forms at once. A charter is a physical archival object with a shelfmark and material properties; it may have one or more digital images, a catalogue record, a transcription, a *regest*, a printed edition and relations to people, places or events. Each representation has its own provenance and may contain information absent from the others. Treating all of them as a single record obscures these distinctions, whereas separating them completely makes navigation difficult. Digital Livonia addresses the problem through relational structures that connect records while retaining their separate functions. Controlled vocabularies are used where stable comparison is required, although the treatment of variant medieval names and uncertain identifications remains uneven across collections. A fuller CIDOC CRM (Conceptual Reference Model) mapping would make the meanings of these relations more explicit beyond the local database. Its implementation requires each collection's categories to be examined, ambiguous relations to be represented and the provenance of imported information to be retained.[68] Our immediate priority has been to publish usable collections and establish selected connections between them. Further semantic alignment can proceed within this architecture, provided that the underlying records preserve the distinctions needed for subsequent mapping.

Images are handled through the International Image Interoperability Framework (IIIF). IIIF separates the delivery of an image from the application in which it is viewed. Its Image API allows a client to request a specified region, size, rotation, quality and format through a standardised address, while the Presentation API describes the sequence and structure of compound digital objects such as codices, registers or multi-page documents. Textual annotations associated with these objects can be searched through the Content Search API.[69] The practical advantage is that an image held by one institution can be displayed,

68 C. Bekiari et al. Definition of the CIDOC Conceptual Reference Model. Version 7.1.3. February 2024. https://cidoc-crm.org/sites/default/files/cidoc_crm_version_7.1.3.pdf (accessed 20 September 2026).

69 International Image Interoperability Framework Consortium. IIIF Image API 3.0, 2020, https://iiif.io/api/image/3.0/; IIIF Presentation API 3.0, 2020. https://iiif.io/api/presentation/3.0/; IIIF Content Search API 2.0, 2022. https://iiif.io/api/search/2.0/ (accessed 20 September 2026).

compared or annotated in another compatible environment without transferring the master file. For Digital Livonia, IIIF delivery draws on two sources. Printed source editions held in the Tallinn University Academic Library's ETERA repository are served through ETERA's own IIIF output, which supplies a durable institutional setting for this material. For images managed directly within Directus, the project assembled its own image-processing and delivery pipeline: a self-contained Docker service using IIPImage.[70] This supports local and S3-compatible storage, provides a IIIF Image API 3.0 endpoint and offers a REST interface for transcoding images. Since the service is independent of a particular content-management system, it can continue to deliver images if the Digital Livonia frontend or data-management layer changes. Two Directus extensions were developed to connect this infrastructure to the platform, one generating IIIF Presentation manifests and the other managing image delivery. Their release as open-source components allows the work undertaken for a regional project to be reused elsewhere.[71]

This architecture changes the status of the digital image. In a conventional website, an image is often an illustration embedded in a page and used only within that page. In an IIIF environment, the image becomes an addressable resource that can be cropped, enlarged, cited and brought into a different viewer. A manuscript page held by the Tallinn City Archives can be displayed beside a related fragment or parallel text served by another institution, provided both use the same standards. The larger promise is the virtual reconstruction of documentary contexts dispersed by the history of preservation. As the preceding discussion has emphasised, such reconstruction cannot restore the material object in all its dimensions, but it can create forms of comparison that would otherwise require repeated travel and the simultaneous cooperation of several holding institutions.

Transcription constitutes a further layer. Project texts are prepared in Transkribus, where document images can be segmented, transcribed, corrected and enriched with structural or textual tags. Transkribus permits export in formats ranging from plain text and searchable PDF to PAGE XML, ALTO XML, spreadsheets and TEI XML.[72] In the Digital

70 IIPImage is an advanced high-performance image server system for web-based streamed viewing and zooming of ultra high-resolution images, https://iipimage.sourceforge.io/ (accessed 20 September 2026).

71 IIIF Conversion and Image API End-point. GitHub repository, https://github.com/digital-livonia/iiif-convert-and-serve/ (accessed 20 September 2026).

72 Transkribus. Downloading: Formats Available to Store, Publish or Further Analyse Transcriptions. – Transkribus Help Center, https://help.transkribus.org/downloading/ (accessed 20 September 2026).

Livonia workflow, we convert ALTO XML into IIIF annotations, so that the resulting transcription is aligned with the corresponding images and prepared for IIIF presentation and content search. This procedure allows the manuscript or document to remain visible while its text becomes searchable and reusable. It also avoids making Transkribus itself the final publication environment. The transcription can move from the recognition and editing service into the platform, a digital edition or another analytical workflow.

Search operates across several technical layers. The platform queries its own structured datasets and selected content preserved in ETERA, the digital repository of the Tallinn University Academic Library. Through IIIF Content Search, annotations associated with the library's materials and texts transcribed within the project can also be searched. The original proposal envisaged a semantic search that would allow users to specify whether they were looking for people, places, events, institutions or other relations connected with a search term. Elements of this entity-centred approach have been implemented in the current people and places interfaces, but a uniform semantic search across all collections is still under development. Full-text search, field-based database search and semantic search operate at different analytical levels and therefore answer different questions. Finding the occurrence of a name in a transcription does not establish which historical person it denotes, whereas retrieving all records linked to an identified individual depends on prior scholarly work of disambiguation.

The public interface and the principal collection descriptions are available in Estonian and English, while the sources retain their historical languages and record-level translation remains uneven. Multilinguality therefore extends beyond the interface itself. The longer-term aim is to reconcile Latin, Middle Low German and vernacular forms with modern Estonian, Latvian, German, and English names without obscuring the form attested in the source. Variant labels are currently preserved within individual datasets, but a platform-wide multilingual authority layer has not yet been established. Developing such a layer will require stable entity identifiers, language-tagged labels, explicitly defined preferred and alternative names, and documentation of uncertain identifications.

The architecture is organised as a connected sequence of services rather than as a single self-contained system. Institutional repositories preserve and serve digital objects; IIIF standardises their delivery and presentation; Transkribus supports transcription; Directus manages structured research data and exposes them through APIs; the frontend

brings these services together for the user; and downloadable data can be transferred to GIS, network-analysis or statistical applications. In this respect, the architecture resembles the 'end-to-end' infrastructure described by Waters, although not every link in the chain has yet been completed and several transitions still depend on substantial manual mediation.[73] Its coherence lies in the movement of materials between services and in the standards that make this movement possible.

RESEARCHER WORKFLOWS AND THE HUMAN INFRASTRUCTURE OF DATA

The most instructive problems encountered during construction concerned neither APIs nor image servers, but the relationship between the data model and the working habits of researchers. The project team consisted primarily of historians, alongside art historians and archaeologists whose habitual tools were Word, Excel, personal notes and, in some cases, handwritten tables. Their expertise lay in reading difficult sources, identifying people and places, reconstructing chronologies and evaluating uncertain evidence rather than creating a database or a web application whose value depends on discrete, reusable and consistently structured records. They did not join the project in order to learn the internal logic of a complex database interface, and it gradually became clear that expecting them to adapt their work to such an interface would consume time better devoted to the sources.

The initial design assumed that researchers would enter information directly into forms reflecting the intended relational structure. In practice, some found the interface difficult and postponed data entry, while others used free-text fields or improvised conventions that circumvented the model. The resulting problem was not a lack of scholarly competence. The interface had exposed the technical structure at the point where researchers needed intellectual flexibility. A historian gathering information from an account book often does not know at the beginning which categories will prove significant, whether two names denote the same person, or how a doubtful date should eventually be represented. Premature formalisation can interrupt this exploratory process and encourage uncertain judgements to be entered as stable facts.

73 D. J. Waters. The Emerging Digital Infrastructure for Research in the Humanities, 88–92.

A more workable procedure began from the tools that researchers already used, especially spreadsheets. These already impose a model through their columns, rows and recording conventions, but allow that model to be revised during exploratory work. Their limitations became apparent when the data had to represent multiple relations, maintain consistent values or support collaborative revision. We increasingly used spreadsheets as provisional research environments from which records could be imported into relational collections through scripts and agreed procedures. This shifted the point at which stricter validation was required, while leaving historians and technical staff jointly responsible for checking the transformations. For this project, accommodating familiar working practices proved more effective than requiring researchers to adopt the database interface at the outset. It did not remove the need for common conventions or changes in how evidence was recorded. It also transferred work to those preparing imports, reconciling categories and checking transformed records. The arrangement therefore depends on procedures through which researchers can review those changes and return corrections to the maintained dataset. Our experience suggests that the timing of formalisation needs to be negotiated alongside responsibility for the work it creates.

The complexity of historical evidence need not be incompatible with the requirements of structured data. Organising evidence as data necessarily involves simplification, but it should preserve the distinctions on which historical interpretation depends.[74] A date may be exact, approximate or established within limits defined by two documented events; a place name may have several plausible identifications; and a person may appear under different names or be difficult to distinguish from a namesake. Absence from a source may likewise reflect a gap in the record rather than evidence of non-existence. Database fields need to accommodate these uncertainties without requiring researchers to claim greater precision than the evidence allows. The design of Digital Livonia benefited from a growing emphasis on how the system could represent the state of historical knowledge, including its uncertainties, which prompted reconsideration of how sources were accommodated within it.

This process also made visible the human infrastructure behind the platform. Every record depends on archival identification, transcription, normalisation, checking and decisions about correspondence between

74 J. Drucker. Humanities Approaches to Graphical Display. – Digital Humanities Quarterly, 2011, 5, 1. https://dhq-static.digitalhumanities.org/pdf/000091.pdf.

sources. Technical specialists must understand enough of the historical problem to avoid building formal relations that distort the evidence; historians must understand enough of the data model to recognise what their classifications will permit or exclude. Librarians and archivists contribute knowledge of collections, identifiers, rights and preservation. Open-source communities maintain the standards and software on which the platform relies. The finished interface conceals much of this labour, but its reliability is inseparable from it. Digital infrastructure consists as much of expertise, conventions, documentation and institutional trust as of code and servers.

DIGITAL LIVONIA AMONG MEDIEVAL RESEARCH PLATFORMS

Digital Livonia belongs to a wider ecology of portals and research environments created during the past three decades. We compare Digital Livonia with selected projects to clarify how its scale, architecture and scholarly functions relate to other approaches. Manuscriptorium, maintained by the National Library of the Czech Republic, represents the model of a large international aggregator. It brings together catalogue records and digitised documents from hundreds of institutions and supports interoperable access through IIIF.[75] Digital Livonia operates on a much smaller regional scale and combines aggregated material with databases and editions created through the project's own research; its emphasis therefore falls less on the breadth of discovery than on establishing relations between collections and historical entities.

Monasterium.net and Fragmentarium illustrate research environments organised around particular source types. Monasterium's collaborative archive provides access to more than half a million charters and related documents organised into archival fonds and research collections. Standardised metadata, the Charters Encoding Initiative and mechanisms for scholarly contribution allow documents from different institutions to be described, edited and reused within a common environment.[76]

75 Manuscriptorium: Gateway to Written Cultural Heritage, https://www.manuscriptorium.com/en/front-page-english/ (accessed 20 September 2026). On its new full-text edition module, see Manuscriptorium. Full-Text Edition Module, 29 May 2026, https://www.manuscriptorium.com/en/updates/-/full-text-edition-module/ (accessed 20 September 2026).

76 Monasterium.net: The Collaborative Archive, https://www.monasterium.net/mom/home/; Monasterium.net, Import Charters, https://www.monasterium.net/mom/import-charters/ (accessed 20 September 2026).

Fragmentarium addresses the particular problems posed by medieval manuscript fragments, supporting cataloguing, tagging, transcription and the virtual reconstruction of dismembered codices, with IIIF enabling images from partner institutions to be brought into a shared working environment.[77] Their concentration on particular source types supports specialised operations and shared descriptive conventions. Digital Livonia's regional remit requires it to connect more heterogeneous evidence, but both projects offer useful precedents for developing research functions within individual collections. Monasterium's collaborative architecture also suggests how controlled contributions, transparent revision histories and procedures for returning enriched information to holding institutions could support a continuing scholarly workspace.

Biblissima and the Finnish Sampo portals provide the closest comparisons for Digital Livonia's longer-term entity-centred ambitions. Biblissima aggregates catalogues and research databases concerning the transmission of texts and images from antiquity to the Renaissance, organising access around authority-controlled entities such as manuscripts, works, people, places, institutions and collections. Shared identifiers and IIIF services connect resources that retain their distinct institutional origins.[78] The Sampo model separates reusable linked-data services from the portals through which their contents are searched, visualised and analysed, using shared entities and semantic relations to connect heterogeneous datasets.[79] Mapping Manuscript Migrations offers a medieval example: it harmonises metadata on more than 220,000 medieval and Renaissance manuscripts, supporting the investigation of manuscripts, provenance events, owners and places through faceted search, analytical views and a reusable SPARQL endpoint.[80] These projects demonstrate the editorial and institutional investment required to connect independently constructed resources. Their scale exceeds Digital Livonia's present capacities, but their architecture provides a precedent for selectively aligning entities across its collections.

77 Fragmentarium. Digital Laboratory, https://fragmentarium.ms/pages/about/digital-laboratory/ (accessed 20 September 2026).

78 Biblissima: Observatory of Written Cultures, https://portail.biblissima.fr/en/; Biblissima, Data Sources, https://portail.biblissima.fr/en/data-sources/ (accessed 20 September 2026).

79 E. Hyvönen. Digital Humanities on the Semantic Web: Sampo Model and Portal Series. – Semantic Web, 2023, 14, 4, 729–744. https://doi.org/10.3233/SW-223034; Sampo Model, Data Services, and Series of Semantic Portals. Semantic Computing Research Group, Aalto University, https://seco.cs.aalto.fi/applications/sampo/ (accessed 20 September 2026).

80 E. Hyvönen, E. Ikkala, M. Koho, J. Tuominen, T. Burrows, L. Ransom, H. Wijsman. Mapping Manuscript Migrations on the Semantic Web: A Semantic Portal and Linked Open Data Service for Premodern Manuscript Research. – The Semantic Web – ISWC 2021. Ed. by A. Hotho et al. Springer, Cham, 2021, 615–630. https://doi.org/10.1007/978-3-030-88361-4_36.

People of Medieval Scotland offers an equally pertinent comparison from digital prosopography. Its factoid-based model represents statements about people, offices, places and relationships as source-specific assertions rather than merging them silently into a single modern biography. Conflicting or uncertain information can therefore remain attached to the evidence from which it derives, while the aggregated data can support maps, network analysis and linked-data reuse.[81] Its regional focus and source-centred treatment of identity are especially relevant to Digital Livonia, where people must frequently be reconstructed from variant names and fragmentary attestations distributed across independently created databases.

For Digital Livonia, these precedents suggest developing specialised functions within collections while connecting selected entities across them. The extent of integration will depend on the scholarly value of each connection and the editorial work required to establish and maintain it. The comparison also warns against assuming that digital availability is tied permanently to a particular technical implementation. The Digitized Medieval Manuscripts app (DMMapp), a widely used directory linking users to digitised manuscript collections, has announced that its current website will cease operation on 22 November 2026, although a successor version of the application has already been made available.[82] This case illustrates both the vulnerability and the potential resilience of digital scholarly resources: hosting costs and maintenance demands can render one implementation unsustainable, while open code, reusable data and careful documentation can facilitate migration to a new environment. Interoperable standards and open-source development therefore improve the prospects of continuity, although long-term sustainability still depends on the availability of technical maintenance, governance and institutional or community support.

81 A. Beam, J. Bradley, D. Broun, J. R. Davies, M. Hammond, N. Jakeman, M. Pasin, A. Taylor et al. People of Medieval Scotland: 1093–1371. Glasgow, London, 2019, https://poms.ac.uk/ (accessed 4 August 2026); M. Pasin, J. Bradley. Factoid-based Prosopography and Computer Ontologies: Towards an Integrated Approach. – Digital Scholarship in the Humanities, 2015, 30, 1, 86–97. https://doi.org/10.1093/llc/fqt037.

82 DMMapp: Digitized Medieval Manuscripts App, current site (https://digitizedmedievalmanuscripts.org/) and successor application (https://sexycodicology.github.io/Digitized-Medieval-Manuscripts-app/) (accessed 20 September 2026). The existing website announces that it will cease operation on 22 November 2026 and directs users to a newly available version of the application. See also G. Menna. The End of an Era: Saying Goodbye to the DMMapp after 12+ Years. – DMMapp Blog, 19 December 2025. https://blog.digitizedmedievalmanuscripts.org/the-end-of-an-era-saying-goodbye-to-the-dmmapp-after-12-years/ (accessed 20 September 2026).

LIMITS AND FUTURE DEVELOPMENT

The Digital Livonia platform now constitutes the first stable configuration of infrastructure that remains under development. Several of its present limitations reflect the history of its construction. Metadata varies in scope and granularity because the collections were assembled from sources created at different times and for different purposes. Some databases are comprehensive within a narrowly defined corpus, whereas others remain preliminary and explicitly anticipate the addition of further records. Relations between people, places, documents and objects have been established in selected collections but have not yet been normalised across the platform as a whole. A uniform semantic search, comprehensive CIDOC CRM mapping, and general indexes of events and institutions remain under development.

Evaluation has so far been formative rather than systematic. Feedback from project researchers shaped data-entry interfaces and workflows during development, but no published user study, citation analysis or public usage report is yet available. The next phase is intended to combine task-based testing with historians and collection custodians, structured feedback on failed searches and ambiguous records, and API and download metrics. Such measures would assess how effectively the platform supports reproducible research tasks rather than merely recording the number of visits to its pages.

These limitations are also methodologically revealing. A regional knowledge infrastructure does not become comprehensive simply through the accumulation of records. Its usefulness depends on decisions about identity, provenance, uncertainty and the meaning of relations. Before two occurrences of a name can be linked to a single person, their dates, offices, social connections and documentary contexts must be evaluated. Before a historical place can be mapped, variant names and possible locations must be reconciled without concealing doubt. Before data imported from another repository can be enriched, the original description must be distinguished from additions made within the project, and responsibility for each assertion must remain identifiable. The slow work of entity resolution and provenance is therefore central to the platform's further development.

The next phase will concentrate on connections that can be established and evaluated within the available resources. Selected people, places and institutions recurring across collections provide a starting point, but their identification must retain the supporting evidence and

any unresolved alternatives. Documentation of schemas, vocabularies, provenance and revision procedures is needed alongside this work so that exported records remain intelligible outside the platform. Multilingual labels and external authority links can then be developed without suppressing local forms. These improvements also require explicit responsibility for data correction, versioned releases, preservation and software maintenance beyond individual research grants. The FAIR principles provide a useful orientation for this work, since the platform aims to make its data findable, accessible, interoperable and reusable. Historical data complicate each of these terms. Findability depends on names that vary across Latin, Middle Low German, Estonian, Latvian and modern scholarly usage. Accessibility may be constrained by the rights attached to images held by different institutions. Interoperability requires mappings that preserve rather than erase differences between source traditions. Reuse depends on documentation sufficient to explain how a modern record was derived from medieval evidence. Maintaining these qualities requires continued work as records are corrected, collections expand and the platform changes.

Following the completion of the original research grant at the end of 2025, the platform has remained hosted by ETAIS, the Estonian Scientific Computing Infrastructure, operated as shared national infrastructure under the Estonian Research Infrastructures Roadmap, while the Academic Library continues to provide the ETERA repository and image-delivery layer. Its further development is currently supported by 'Digital Livonia 2.0: For a Digitally Enhanced Study of Medieval Livonia (c. 1200–1550)' (TUF25012), funded by Tallinn University. Its broader institutional setting now includes Tallinn University's participation in the Estonian Research and Cultural Data Infrastructure for 2025–2029. Further continuity will be provided by a new five-year research project funded by the Estonian Research Council, due to begin in early 2027 and devoted to the study of medieval Livonian sources in the Vatican Apostolic Archives through methods and services supported by the Digital Livonia infrastructure. Together, these arrangements strengthen the local environment in which the platform can be maintained and further developed, although a platform-specific governance policy still needs to be formalised.

Sustainability also requires a clear distinction between components that must be preserved and those that can be replaced. The public interface will age and should be expected to change. The underlying data, stable identifiers, documentation and links to source images have a longer

scholarly life. IIIF enables images to remain usable in other viewers; APIs allow structured data to be consumed by new applications; and institutional repositories provide a more durable home for master files than a project server. The modular architecture of Digital Livonia was designed with this distinction in mind. The longer-term measure of its success will be whether future researchers can understand, cite and reuse its materials after the present combination of software has ceased to be current.

CONCLUSION

The developments surveyed in this article show how digital methods have expanded access to medieval sources and altered the scale and organisation of their comparison. Their consequences extend into the constitution of evidence itself: imaging determines which material features become visible, transcription and encoding select what can be searched, and computational models establish the terms on which patterns are identified. Across these operations, source criticism must address both the historical formation of the surviving material and the procedures through which it becomes available for analysis. The history of digital medieval studies also shows that these questions have accompanied computer-assisted research from its beginnings, even as its instruments and institutional settings have changed.

The construction of Digital Livonia brings these methodological questions into the practical work of organising a regional research infrastructure. Our initial ambition for comprehensive semantic integration encountered differences between sources, datasets and researchers' working practices that could not readily be resolved within a common model. An incremental architecture allowed collections to become usable while their relationships remained under development. Accommodating familiar tools such as spreadsheets helped researchers contribute their material, but transferred further work to those preparing, checking and maintaining the data. These choices enabled publication and reuse while leaving substantial questions of entity identification, provenance, multilingual access and institutional responsibility unresolved.

Read together, the survey and the platform study suggest that a hybrid practice of medieval research depends on sustained engagement between digital procedures and philological, palaeographical,

codicological and historical judgement. Such engagement begins when sources are selected and represented and continues through modelling, analysis and publication. For regional infrastructures such as Digital Livonia, its effectiveness depends on whether researchers can follow the connections between sources and scholarly assertions, recognise uncertainty and revise earlier interpretations. Their lasting contribution will rest on keeping those connections intelligible and open to scrutiny as research questions, software and institutional arrangements change.

Acknowledgements

The research and writing of this article were supported by two projects devoted to the development of digital approaches to medieval Livonia: Digital Livonia: For a Digitally Enhanced Study of Medieval Livonia (c. 1200–1550) (PRG1276), funded by the Estonian Research Council, and Digital Livonia 2.0: For a Digitally Enhanced Study of Medieval Livonia (c. 1200–1550) (TUF25012), funded by Tallinn University.

DIGIPÖÖRE KESKAJAUURINGUTES: MEETODID, TARISTUD JA „DIGITAALSE LIIVIMAA“ KOGEMUS

Marek Tamm, Hembo Pagi

Digipööre on viimase paarikümne aasta jooksul muutnud keskajauuringute peaaegu kõiki tööetappe, alates allikate leidmisest ja transkribeerimisest kuni nende analüüsimise, tõlgendamise ja publitseerimiseni. Digitaalsed keskajauuringud ei ole seetõttu enam üksnes kitsas eriala digihumanitaaria sees, vaid üha enam keskajauuringute üldine töökeskkond. Samas ei ole tegemist katkestusega senises uurimistraditsioonis. Digitaalsed meetodid on kujunenud pika ja ebaühtlase arengu tulemusena ning eksisteerivad kõrvuti trükikataloogide, kriitiliste väljaannete, arhiivitöö ja lähilugemisega. Artiklis eristatakse analüütiliselt kolme kattuvat tegevust: 1) digiteerimine teeb allikad ja uurimistulemused elektrooniliselt kättesaadavaks; 2) digitaalne struktureerimine võimaldab neid otsida, seostada ja taaskasutada; 3) arvutuslik analüüs rakendab tekstide, piltide ja ajalooliste andmete uurimisel statistilisi või algoritmilisi meetodeid.

Digitaalsete keskajauuringute kujunemislugu ulatub Teise maailmasõja järgsesse aega. Üheks selle sümboolseks alguspunktiks on Roberto Busa „Index Thomisticus", kuid oluline on silmas pidada, et arvutipõhiste keskajauuringute teke oli märksa mitmekesisem protsess. 1960.–1980. aastatel arendati mitmes Euroopa uurimiskeskuses keskaegsete ladina tekstide automaatseid konkordantse, ajaloolisi andmebaase ja leksikomeetrilisi korpusi. 1990. aastatel avasid personaalarvutid, internet ja teksti kodeerimise standardid uue etapi, mida esindasid näiteks projektid „Electronic Beowulf" ja „Piers Plowman Electronic Archive". Alates 2010. aastatest on esiplaanile tõusnud omavahel ühendatud töövood, kus digiteerimine, transkribeerimine, annoteerimine, analüüs ja publitseerimine moodustavad üha enam ühe tervikliku teadustaristu.

Digitaalsete kogude suurim vahetu mõju on olnud ligipääsu laienemine. Rahvusvaheline pildikoostalitlusraamistik IIIF võimaldab eri mäluasutustes hoitavaid käsikirju ühises keskkonnas vaadelda, võrrelda ja annoteerida. Digiteerimine ei kõrvalda siiski vajadust allikakriitika järele. Veebis kättesaadav materjal on paratamatult valikuline ja digitaalne kujutis muudab materiaalse objekti omadusi. Teisalt võivad multispektraalne fotograafia, pinnastruktuuri analüüs ja 3D-modelleerimine tuua nähtavale selliseid tunnuseid, mida palja silmaga on raske või võimatu märgata. Digitaalne allikakriitika peab seetõttu küsima nii kujutise loomise, metaandmete päritolu ja valikute kui ka identifikaatorite, kasutusõiguste ja säilitamise kohta.

Oluline osa digitaalsetest keskajauuringutest on seotud kriitiliste editsioonide ja käsikirjade masinloetavaks muutmisega. Digitaalne väljaanne ei ole pelgalt mahukam trükiväljaanne, vaid tekstipärimuse eksplitsiitne mudel, milles tuleb määratleda tekstiüksused ning lühendite, paranduste, ebakindlate lugemite ja eri käsikirjade omavahelised suhted. Käsikirjalise teksti automaattuvastus ehk HTR on omakorda märgatavalt laiendanud võimalusi suurte käsikirjakorpuste transkribeerimiseks, samas kujundavad transkriptsioonireeglid juba ette seda, millised käsikirja tunnused muutuvad hiljem arvutuslikult analüüsitavaks. Normaliseerimine võib lihtsustada teksti lugemist ja otsimist, kuid kaotada paleograafiliselt olulist teavet. Seetõttu on HTR kõige tulemuslikum inimese ja masina koostöös.

Masinloetavad tekstid võimaldavad uurida keskaja tekstikorpusi sellises ulatuses, mida klassikaline lähilugemine ei võimalda. Stilomeetria, autorluse määramine, teksti taaskasutuse tuvastamine, semantilised mudelid ja muud arvutuslikud meetodid toovad nähtavale tekstimustrid, mis on hajutatud sadade või tuhandete käsikirjade vahel. Nende

tulemused vajavad siiski hoolikat tõlgendust. Statistiline sarnasus võib tuleneda autorist, žanrist, kirjutajast, ümberkirjutamisest, toimetamisest või transkriptsioonivigadest. Sama kehtib võrgustikuanalüüsi ja ajaloolise GIS-i kohta. Võrgugraafi seos on juba tõlgendus: sugulus, kirjavahetus, üriku kaastunnistamine või inkvisitsiooniprotokollis mainimine ei ole samalaadsed suhted. Ka ajaloolise ruumi sidumine tänapäevaste koordinaatidega võib luua eksitava täpsuse. Hea andmemudel peab seetõttu säilitama ebakindluse ja võimaldama konkureerivaid tuvastusi.

Generatiivse tehisaru levik on toonud keskajauuringutesse uusi võimalusi ja riske. Suured keelemudelid võivad toetada otsingut, tõlkimist, infoeraldust ja töövahendite loomist, kuid väljundi ladusus ei ole usaldusväärsuse mõõt. Keskajal kasutatud keeled ja tekstipärimused on treeningandmetes ebaühtlaselt esindatud ning kommertsmudelite puhul jäävad sageli varjatuks kasutatud väljaanded, normaliseerimispõhimõtted ja mudeli täpne päritolu. Artiklis käsitletakse ühe katsetamisvõimalusena kureeritud korpustel põhinevat infootsingut, mis seob genereeritud vastuse kontrollitavate tekstikohtadega. Ka sellisel juhul tuleb eraldi hinnata leitud tõendite asjakohasust ja vastuse täpsust, sest viidatud tekstikoht ei pruugi väidet toetada ning oluline teave võib jääda leidmata. „Digitaalse Liivimaa" avalikku kasutajaliidesesse ei ole üldotstarbelist vestluslikku tehisaru seni lisatud.

Artikli teine osa analüüsib „Digitaalse Liivimaa" veebiplatvormi kui piirkondliku teadustaristu näidet. 2021. aastal Tallinna Ülikoolis algatatud projekti eesmärk on koondada ja seostada keskaja Liivimaa uurimiseks vajalikke hajutatud allikaid ja uurimisandmeid. Liivimaa on selleks eriti sobiv näide, sest ajalooline piirkond ei lange kokku ühegi tänapäevase riigi territooriumiga ning selle allikad paiknevad Eesti ja Läti kõrval Saksamaa, Vatikani, Poola, Rootsi, Taani, Venemaa ja teiste maade kogudes. „Digitaalse Liivimaa" avalik keskkond koondab digiteeritud käsikirju, digitaalseid allikaväljaandeid, materiaalseid objekte, kolmteist temaatilist andmebaasi ja kaks arvutuslikku rakendust. Andmebaasid hõlmavad muu hulgas kinnisvara, vaimulikke, Saksa ordu Liivimaa haru rüütelvendi, raamatuid, käsikirjakoodekseid, mõisaid, hindu, palku ja Eesti keskaegsete väikelinnade elanikke. Rahaühikute- ja kaalukalkulaator näitavad, kuidas ajalooliselt muutlikke mõõtühikuid saab formaliseerida ajaliselt ja piirkondlikult täpsete reeglite alusel. Platvormi arhitektuur on moodulipõhine: institutsionaalsed hoidlad säilitavad algfaile, IIIF tagab piltide standardiseeritud esituse, Transkribus toetab transkribeerimist, Directus haldab struktureeritud andmeid ning API-liidesed võimaldavad neid kasutada ka väliste GIS-i, võrgustiku- ja statistiliste töövahenditega.

Projekti üks olulisemaid kogemusi puudutab uurijate tööpraktikat. Algne eeldus, et ajaloolased sisestavad andmed otse keerukasse relatsioonilisse andmebaasi, osutus liiga piiravaks. Tõhusamaks kujunes töövoog, kus uurijad said esialgu kasutada tuttavaid vahendeid, eeskätt arvutustabeleid, ning andmed formaliseeriti hiljem tehniliste protseduuride ja spetsialistide toel. Sellest tuleneb üks projekti keskseid põhimõtteid: tööriistad tuleb kohandada uurijate tööviisidele enne, kui uurijatelt nõutakse kohanemist tööriistadega. Digitaalne taristu koosneb seega samavõrd inimestest, teadmistest, kokkulepetest ja institutsionaalsest vastutusest kui tarkvarast ja serveritest.

„Digitaalse Liivimaa" kogemus näitab, et piirkondliku teadustaristu loomisel tuleb otsustada, milliseid andmeid ja seoseid saab ühtlustada ning millised allikate ja uurimisküsimuste erinevused peavad säilima. Platvormi praegune sisu on geograafiliselt ebaühtlane ning kogude omavaheline sidumine alles jätkub. Edasise arenduse keskmes on valitud isikute, paikade ja institutsioonide tuvastamine eri andmekogudes, tõlgendusotsuste dokumenteerimine ning uurimisandmete säilitamise ja ajakohastamise vastutuse määratlemine.