

A DIGITAL ANALYSIS OF ARTISANS' STATUTES – AN INTRODUCTION OF DIGITAL METHODS IN MEDIEVAL LIVONIAN HISTORY RESEARCH

RŪTA BRUŽEVICA

Mg. hist., PhD student and junior research fellow

at Tallinn University School of Humanities

E-mails: ruta.bruzevica@tlu.ee

ABSTRACT

The field of history is rapidly implementing digital humanities tools in researching various historical periods and questions. In medieval history research, one of the major technologies to help with the work with sources is Handwritten Text Recognition, allowing a faster transcription of sources. Source digitalisation results in digital editions and datasets, providing sustainable research data for open and future use. This article lists several digital tools suitable for use in researching medieval Livonian history with an example of Artisans' Statutes in the process of full-scale digitalisation.

Keywords: digital history, Digital Livonia, Artisan Statutes, Transkribus, Nodegoat.

INTRODUCTION

For the past few decades, the digital possibilities provided by computer technologies have affected the daily work of millions of people, and researchers of the Humanities are no exception. While broadly used tools like the Internet, MC Office, and others are common to most, there are also tools created especially for research in the humanities, presented by the field of digital humanities. The methods of digital humanities have been broadly acquired by researchers in various disciplines, especially linguistics, literature, politics, and history. As the field grows, practices of using digital tools for researching history have now been put under a narrower field

known as digital history, as well as digital medieval history for research of medieval history.

Despite the new approaches, the sources used by medieval historians have largely not changed and it is often the reason not to take advantage of the digital methods. However, recently there have been good examples of digital tools that help in researching medieval documents, for example, Handwritten Text Recognition (HTR) technology, practice of publishing digital editions, storing sources and their metadata to provide automatic data analysis, and platforms for creating comprehensive visualisations.

Some of these digital methods are applied by researchers working on the project “Digital Livonia” (Digital Livonia) carried out in Tallinn University since 2021. The outcome of the project will be digital editions of sources, databases, and a searchable portal for history of medieval Livonia. A contribution to the portal will be the work done prior to a PhD thesis on sociability of artisans in medieval Livonian towns. It will include a digital edition compiling all existing original, copied, or already edited versions of medieval artisan statutes from Livonian towns, as well as a database with information about organisations, places, dates, persons, and practices mentioned in the statutes. To achieve the research goal and introduce digital methods in research of medieval Livonia, this research will include experiments with various digital methods for research, e.g., Handwritten Text Recognition, content analysis, source editing and coding, data extraction, network analysis, and other digital methods and tools that might become available in the following years. This article will list some of these methods and introduce their functionality when working with sources of medieval Livonia, particularly, artisan statutes from Riga and Reval (Tallinn).

DIGITAL TRANSCRIPTIONS OF MEDIEVAL STATUTES

Digital tools and programmes are being used more and more often among historians, including medievalists. There have even been claims that often medievalists are early adopters of the digital, with examples of electronic medieval source editions available not long after the Web appeared (Unsworth 2012, §18), successfully applying Geographic Information Systems (GIS) and other digital tools in research projects. With rapid development of technology, new tools that are designed especially for the Humanities and History research are being designed and implemented in individual and institutional projects. Such tools will also be implemented in researching medieval Livonian artisans, with the main ones being an HTR platform *Transkribus* (“*Transkribus*”) for transcription of sources, XML (“XML Editor”) markup language and TEI guidelines (“Guidelines – TEI:

Text Encoding Initiative’) for publishing a digital edition, database and visualisation environment Nodegoat (‘nodegoat | About’, n.d.) for storing data, creating databases and producing analytical visualisations.

When considering the keystones of researching medieval history, it’s impossible to imagine it without sources, whether they are originals in archives, later copies, photocopies, or published editions. As dealing with sources often includes a considerable amount of work, various digital tools have been created to support their availability and research. For the last few decades, the main goal of dealing with medieval sources has been their digitisation¹ by creating digital libraries and electronic manuscript facsimiles have been produced in large quantities by various heritage institutions (Birnbaum et al. 2017, 4–5). Although digital facsimiles have been available, there has been a lack of possibilities to secure fast and precise addition of metadata and transcriptions to these sources, because the previously available Optical Text Recognition (OCR) technology has not been entirely useful for generating transcriptions of handwritten text (Birnbaum et al. 2017, 6), therefore transcriptions mainly have been done by employed researchers or volunteers (Terras 2022, 179). Tools like Transkribus that provide the use of recently developed HTR technology are reshaping this field by allowing to digitally generate transcriptions of these previously digitised facsimiles. Transkribus has become an important tool for hundreds of projects in source digitalisation, e.g., the Citizen Science project for transcribing late Hanseatic diets “Hanse.Quellen.Lesen!” (Popken) by FGHO (Forschungsstelle für die Geschichte der Hanse und des Ostseeraums) and “Mapping Medieval Vienna” by the Free University of Berlin in Germany, (Helmchen) as well as three archival collections have been made available through Transkribus by the National Archives of Estonia (‘The full-text search platform is open – Rahvusarhiiv’). Other Institutions actively using the tool are the Rijksmuseum, British Library, University of Cambridge and others, and together in 2020 Transkribus had more than 18’000 unique active users (Terras 2022, 186).

In research of medieval artisan statutes, Transkribus is used to fully digitalise already digitised (photographed or scanned) statutes, mainly to provide digital transcription and annotation. It is designed to support the transcription of handwritten texts, including historical manuscripts, and offers such tools as a layout editor, a text editor, and a possibility to provide metadata, as well as allows use of different languages and scripts. In this case the setting allows to successfully work with texts in Middle Low German and to create specific models for manuscripts in this language.

The process of digitalising² sources can only begin after a successful and high-quality digitisation of the source, which means having good quality photos or scans of the source provided by the archive or photographed personally. If needed, the .jpeg files have to be edited, e.g. snipped or

enchanted prior to upload, as Transkribus does not offer photo editing on the platform. In this case, the statutes are snipped to a scale 3:4, if their size allows, and enchanted with other editing tools only if they are damaged or with low light.

Once the images are edited and uploaded, it is possible to start creating the transcription. A layout editor allows to define the structure of the document, such as the text regions, text lines, number of columns, notes on sides or other structural specifications. A tool for Layout analysis finishes this process automatically by finding text regions and lines of the manuscript automatically and allowing to edit it, if necessary. This feature is especially useful in such standardised manuscript format as statutes, where text is mainly written in monolithic paragraphs and clear lines, with only a few notes on the sides that sometimes can be easily integrated in the main text region. Creation of a layout and text lines elevates the quality of the transcription, as it is digitally linked to its precise location on the facsimile, as it can be seen on Fig. 1.

When beginning the transcription work on a project, it is not always possible to produce transcriptions by using Artificial Intelligence (AI) only – transcriptions of at least several pages must be done manually if no suitable recognition model is yet available. To make a private model based on the medieval artisan statutes, several transcribed statute manuscripts from 14th to 16th century Tallinn and Riga were trained in an AI model, with a training set of 16'915 words with a resulting Character Error Rate (CER) of 9.27 % on Validation Set (see Fig. 2). CER measures the accuracy of the model, here meaning that the model is accurate 90,73 % of the time, which is a rather good result for a private model, but not effective enough for significant time saving on a transcription.

Training one's own model is not always necessary – there are more than 95 publicly available models on the Transkribus platform, created by researchers or as results of private or institutional projects. Several of those are suitable for automated transcriptions of medieval texts in German, e.g., for 15th century Bastarda, 16th century Current, as well as Gothic book scripts (Public Models in Transkribus). Another public model for 15th–16th century German has been created in 2022 by a group of researchers working on their individual or institutional projects sharing their transcribed documents for a generic model. The model has a geographical scope from Austrian lands to the Baltic Sea, including scripts as Gothic cursive, Bastarda and Early Current (Generic Model 15th–16th century German (prototype)). The artisan statutes of Tallinn and Riga are also among the training data of ~ 77'000 words, contributing to the diverse model and making it as suitable as possible for use on other medieval sources of Livonia. The results of the first generic model created with HTR+

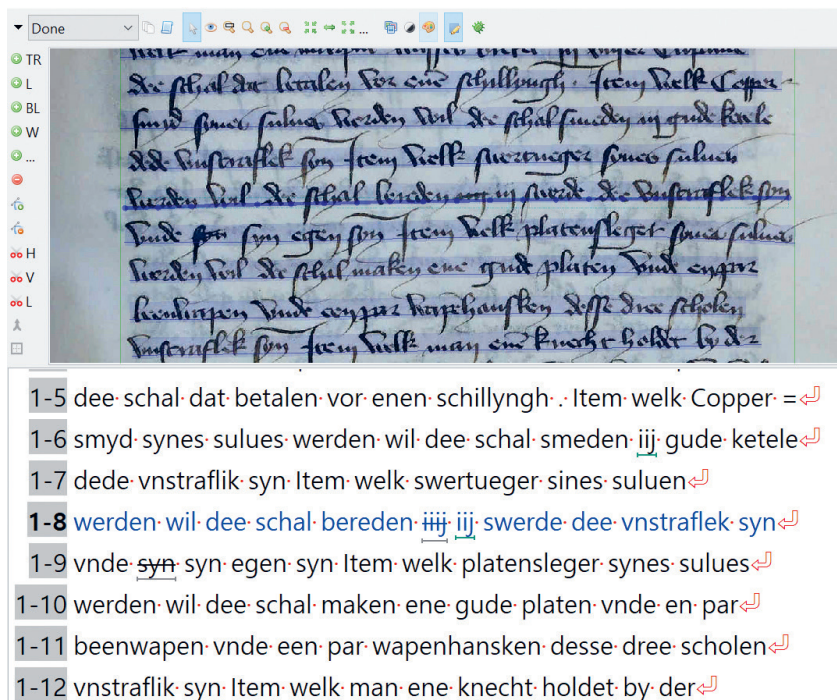


Figure 1. A transcription in progress in the platform Transkribus for the statutes of Riga Blacksmiths from 1382. (LNA-LVVA, 8-1-15, 4r) (photo: R. Bruževica)

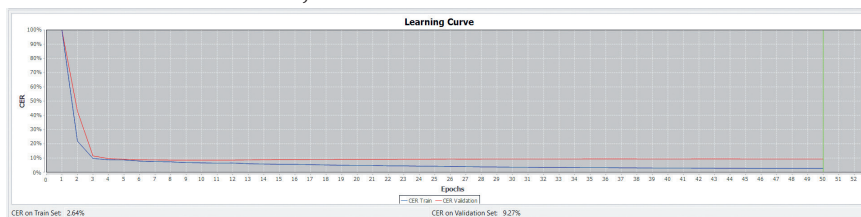
1. attēls. Rīgas kalēju 1382. gada statūtu transkripcijas process platformā Transkribus (LNA-LVVA, 8–1–15, 4r) (autore R. Bruževica)

technology was very successful with a CER of 3.79 %, but the retrained, model with PyLaia technology has a CER of 5.6 %.³ This shows that a model with greater base of material can perform better as it has experience with more diverse handwritings in the same script range. In practice, the generic model also performs increasingly better on other Livonian statutes, than the model based on Livonian manuscripts. Improving the generic model with additional transcriptions might decrease the current CER and make it useful for an easier transcription work and research for other documents created in Livonia during the Middle Ages.

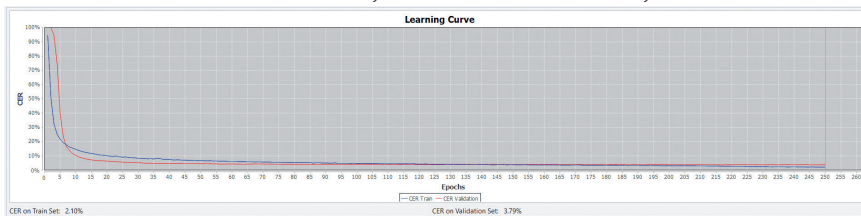
THE IMPORTANCE OF METADATA

Another possibility provided by Transkribus is annotating of the document, its structure and text using metadata. Adding metadata to the transcribed text helps to highlight important facts mentioned in

Medieval Livonia model, CER on validation set 9.27 %



Generic German HTR+ model, CER on validation set 3,79 %



Generic German PyLaia model, CER on validation set 5,60 %

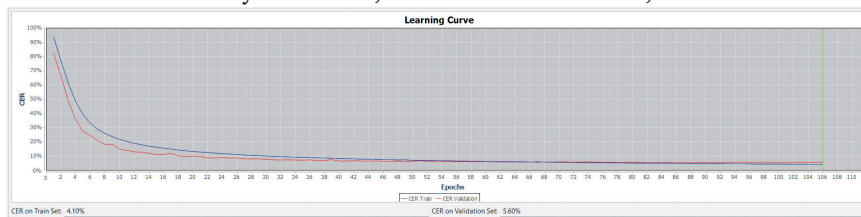


Figure 2. Comparison of Learning curves of three transcription models and their Character Error Rate on Validation set (Graphics by Transkribus, collage by R. Bruževica)

2. attēls. Triju transkripcijas modeļu un to rakstzīmju kļūdu skaita validācijas kopā salīdzinājums (grafiks no *Transkribus*, R. Bruževica)

the text, e.g., date of creation, mentioned places, people, and organisations, as well as to resolve abbreviations and other specifics of the medieval text. In the transcriptions of statutes, the most commonly added text metadata consists of annotations for strikethrough, abbreviations, numbers, dates, places and organisations. Structural metadata is used to distinguish the main paragraph from side notes, drawings or other characters or parts outside of it. At last, document metadata can hold important information about the origin of the digital document, language, time period, or other data desired to describe the digital facsimile of the medieval document.

Although this data is not yet designed to significantly help the transcription work, or design itself automatically, it is important for a later phase of digitalisation – creation of the digital edition. While physical editions have been a solid foundation for making historical sources

accessible, they only allow to represent one layer of information from the source. A digital edition, on the contrary, can hold many, or even all layers of information and it can be improved as often as desirable and necessary (Sahle 2013, 132). It allows the reader to simultaneously access the original manuscript with its original and edited structure, script and language, as well as commentary. Early supplementing of the artisan statute transcriptions with metadata suitable with the widely used XML format and TEI guidelines for publishing digital editions in the humanities, will allow to publish the materials as a digital edition, making it available for sustainable future use and review in the following years. Of course, it is important to publish the edition on a sustainable platform with proper curation and archiving to provide long-term availability (Birnbaum et al. 2017, 12–13).

Further research work with transcriptions and their metadata is possible with other methods, provided by various tools and platforms. Large amounts of collected data is a good ground for various data analysis methods, e.g., social network analysis, geographical analysis, as well as chronological analysis that will be experimented with during the research. One of the tools designed for data analysis in the Humanities, including History, is Nodegoat, a web-based research environment that allows to build an individual data model based on data and research question and offers network, geographical and chronological analysis (Van Bree, Kessels 2017).

The newest features in Nodegoat allow to combine it with the work done in Transkribus – it is possible to ingest finished transcriptions for further adding of the metadata, or creation of nodes in Nodegoat environment for data analysis that is not possible to accomplish using just Transkribus (nodegoat | Guide | Ingest Transcription Data from Transkribus). When researching medieval Livonian artisan statutes, these features will allow to do prosopographical analysis of the artisan practices described in the statutes. An extensive database of metadata defined from work with the statute transcriptions and coding will also allow to conduct chronological analysis on the statutes and their contents to observe how the practices of artisans in Livonian towns have altered over the course of centuries. The data created in Nodegoat is available for extraction to various formats and will stay available for use in future projects on related topics.

CONCLUSIONS

Not every historian can and will implement tools of digital humanities in their research, but those investing their time in gaining knowledge of the field and its possibilities, will profit from it in the upcoming decades. With rapid advancement of AI technologies and their continuous learning

prompted by its users, medieval sources will become more accessible with HTR advancement and might also be applicable to a more precise content analysis. Of course, even AI won't be able to take the place of a historian – machine read sources will always need to be read over by a historian with paleographical knowledge and experience and quantitative analysis will always need an extensive commentary from an expert in the field.

Other advantages that the digital tools offer for research of medieval history, including medieval Livonia, is a possibility to create and later access digital editions with large amounts of data that has already been gathered. Advancement of digital editions and their publications on the history of medieval Livonia, might also contribute to multidisciplinary by creating a common environment for hosting data, analysis and commentary created not only by historians, but also linguists, palaeographers, art historians or others interested in the same source.

Using digital tools for historical sources and data, especially editions and databases, makes the data and research more sustainable and available for the research community, supporting the practice of Open Access in science. Of course, there are challenges to take into consideration, e.g., using sustainable digital formats, sustaining the websites where data is published and ensuring the continuous operation of the hard drives where the data is stored.

Taking in consideration the advantages and possibilities provided by digital tools in the studies exploring various fields of history research, I would like to stress that methods of digital humanities should be implemented in the curricula of history studies to ensure the competitiveness of current students in the developing field of humanities.

ACKNOWLEDGMENT

This work was supported by the Estonian Research Council grant (PRG1276) within the project “Digital Livonia: For a Digitally Enhanced Study of Medieval Livonia (c. 1200–1550)” at Tallinn University.

ABBREVIATIONS

AI – Artificial Intelligence

CER – Character Error Rate

FGHO – Forschungsstelle für die Geschichte der Hanse und des Ostseeraums

GIS – Geographic Information Systems

HTR – Handwritten Text Recognition

LNA-LVVA – Latvian National Archive, Latvian State Historical Archive

OCR – Optical Character Recognition

TEI – Text Encoding Initiative

XML – Extensible Markup Language

BIBLIOGRAPHY

- Digital Livonia – For a Digitally Enhanced Study of Medieval Livonia (c. 1200–1550)*. Available: <https://dl.tlu.ee/> (accessed 18.01.2023).
- Generic Model 15th–16th century German (prototype)*. Available: <https://readcoop.eu/model/generic-model-15th-16th-century-german-prototype/> (accessed 20.01.2023).
- Guidelines – TEI: Text Encoding Initiative*. Available: <https://tei-c.org/guidelines/> (accessed 18.01.2023).
- nodegoat | About*. Available: <https://nodegoat.net/about> (accessed 18.01.2023).
- nodegoat | Guide | Ingest Transcription Data from Transkribus*. Available: <https://nodegoat.net/guide.s/136/ingest-transcription-data-from-transkribus> (accessed 23.01.2023).
- Public Models in Transkribus*. Available: <https://readcoop.eu/transkribus/public-models/> (accessed 20.01.2023).
- The full-text search platform is open! – Rahvusarhiiv*. Available: <https://www.ra.ee/en/the-full-text-search-platform-is-open/> (accessed 18.01.2023).
- Transkribus*. Available: <https://readcoop.eu/transkribus/> (accessed 18.01.2023).
- XML Editor*. Available: https://www.oxygenxml.com/xml_editor.html (accessed 18.01.2023).
- Birnbaum, D. J. et al., 2017. The Digital Middle Ages: An Introduction. *Speculum*. 92, 1–38. Available: <https://www.journals.uchicago.edu/doi/10.1086/694236> (accessed 18.01.2023).
- Van Bree, P., Kessels, G., 2017. ‘nodegoat: Enabling Explorative Research’, 2017 Montral, Canada.
- Helmchen, J. *Mapping Medieval Vienna: The digital edition of historical land registers supported by Transkribus*. Available: <https://readcoop.eu/success-stories/mapping-medieval-vienna/> (accessed 20.01.2023).
- Popken, V. *Rückblicke auf die Niederdeutschen Hansetage: Eine hanseatische Reise mit Transkribus*. Available: <https://readcoop.eu/de/erfolgsgeschichten/recesses-of-low-german-hanseatic-days/> (accessed 19.01.2023).
- Sahle, P., 2013. *Digitale Editionsformen. 2: Befunde, Theorie und Methodik*. Schriften des Instituts für Dokumentologie und Editorik 8. Norderstedt: BoD, Books on Demand.
- Terras, M., 2022. ‘Inviting AI into the Archives: The Reception of Handwritten Recognition Technology into Historical Manuscript Transcription’. In: Jaillant, L., ed. *Archives, Access and Artificial Intelligence: Working with Born-Digital and Digitized Archival Collections*. Digital Humanities Research. 1st edition. Bielefeld: Bielefeld University Press / transcript Verlag. 179–204. Available: <https://www.transcript-open.de/isbn/5584> (accessed 17.01.2023).
- Unsworth, J., 2012. Medievalists as Early Adopters of Information Technology. *Digital Medievalist*. 7. Available: <http://journal.digitalmedievalist.org/article/10.16995/dm.34/> (accessed 18.01.2023).

REFERENCES

- ¹ Digitisation – the process of converting a physical object into a digital format. The result of digitisation is a digital representation of the original document, e.g. a picture or non-searchable pdf file.
- ² Digitalisation – the process of using digital tools and methods to analyse and interpret sources, and to make them more accessible to researchers and the general public.
- ³ Since the end of 2022, Transkribus no longer supports the use of HTR+ and only PyLaiia models are available for use.

AMATNIEKU STATŪTU DIGITĀLA ANALĪZE – IESKATS DIGITĀLAJĀS PIEEJĀS VIDUSLAIKU LIVONIJAS VĒSTURES PĒTNIECĪBĀ

RŪTA BRUŽEVICA

Mg. hist., doktorante, Tallinas Universitātes Humanitārā institūta
jaunākā zinātniskā asistente
E-pasts: ruta.bruzevica@tlu.ee

ANOTĀCIJA

Digitālo humanitāro zinātņu metodes ieņem arvien nozīmīgāku vietu vēstures periodu un jautājumu pētniecībā. Viduslaiku vēstures pētniecībā viena no galvenajām tehnoloģijām, kas palīdz darbā ar avotiem, ir rokrakstu atpazīšana ar mākslīgā intelekta palīdzību, ļaujot ātrāk transkribēt un lasīt vēstures avotus. Avotu digitalizācija, veidojot digitālos avotu krājumus un datubāzes, veicina to ilgtspējību un atvērtu pieeju. Šis raksts apkopo digitālos rīkus un to iespējas viduslaiku Livonijas vēstures pētniecībā, kā piemēru piedāvājot ieskatu viduslaiku amatnieku statūtu digitalizācijā.

Atslēgvārdi: digitālā vēsture, digitālā Livonija, amatnieku statūti, *Transkribus*, *Nodegoat*.

KOPSAVILKUMS

Digitālo humanitāro zinātņu metodes ir plaši apgūtas dažādās disciplinās, tostarp vēsturē. Digitālo rīku izmantošanas prakse vēstures izpētei tagad ir izstrādāta atsevišķa joma – digitālā vēsture, kurā tiek izceltas metodes un rīki, kas piemēroti vēstures pētniecībai. Šis raksts iepazīstina ar pieredzi un iespējām, ko sniedz digitālo humanitāro zinātņu rīki, pētot viduslaiku Livonijas vēsturi, kā piemēru sniedzot amatnieku statūtus. Pilnīgi digitalizējot pieejamos Livonijas viduslaiku amatnieku statūtus, ir iespējams izmantot tādas digitālos rīkus, kas piemēroti viduslaiku pētniecībai, kā *Transkribus* un *Nodegoat*, kā arī metadatu, datubāzu un digitālo avotu krājumu izveide.

Lietotne *Transkribus* ir kļuvusi par noderīgu instrumentu viduslaiku statūtu digitalizācijā, samazinot laiku, kas tiek pavadīts, transkribējot avotus manuāli. Salīdzinot mākslīgā intelekta rokrakstu atpazīšanas modeli, kas balstīts 1) transkribētos Livonijas amatnieku statūtos, ar 2) vispārīgu modeli, kura pamatā ir plašs vēlo viduslaiku dokumentu krājums vācu valodā, pēdējais šajā gadījumā pierādās efektīvāks, kas norāda, ka lielāks

datu korpuss var nodrošināt lielāku precizitāti un līdz ar to jāturpina darbs pie vispārīga modeļa.

Ar *Transkribus* veiktās transkripcijas ir labs pamats digitalizēto avotu turpmākai apstrādei un izmantošanai: 1) veidojot digitālo avotu krājumu, kura pamatā ir humanitārajās zinātnēs bieži izmantotās TEI vadlīnijas, un 2) iespēja iegult transkripcijas lietotnē *Nodegoat*, lai apstrādātu metadatus un veiktu sociālo tīklu, ģeogrāfisko, hronoloģisko un prosopogrāfisko analīzi. Lai gan digitālas metodes un to ieviešana vēstures pētījumos ir izaicinoša, tās paver jaunas durvis manuālā darba samazināšanā, datu uzglabāšanā un avotu ilgtspējā.