

GLAGOLITIC TEXT EDITIONS AND A PERSPECTIVE TOWARDS DIGITAL HUMANITIES

Erich Renhart

*Special Collection Department of Graz
University Library & the Vestigia –
Manuscript Research Centre,
University of Graz, Graz, Austria*

KEYWORDS:

*Digital Humanities, Text Encoding
Initiative, XML, Glagolitic
manuscripts*

ABSTRACT

The Glagolitic written heritage is not yet accessible online or only to a rather limited extent. This article tries to develop a perspective of how to start bridging the gap encouraging the use of extensible markup languages (XML) for cataloguing and text editions. It repeatedly refers to the Comparative Oriental Manuscript Studies Handbook (2015), which reflects the neighbouring ancient text traditions of the Mediterranean region and their achievements in the Digital Humanities.

Proem

It was at the occasion of the “Summer School in the Study of Old Books”¹ held in Zadar in 2009, where I had the privilege to meet Mirna for the first time. I learned a lot from this distinct scholar during the last decennium in terms of her professional competencies – some of which will also become apparent in the article below, and in terms of her charming personality. I became acquainted with a strong character, encouraging and fostering a plethora of young students. Also, once she has established her point of view, Mirna knows how to express her position with conviction. It fills me with pride and pleasure to have been selected to contribute together with a corona of friends and colleagues. What follows is by no means a comprehensive in-depth study, but rather a rough sketch.

Incipit: *Comparative Oriental Manuscript Studies Handbook*

The year 2015 saw the publication of the *COMSt Handbook*, a veritable milestone in the study of oriental manuscripts and the outcome of a five-year programme financed by the European Science Foundation and containing contributions by 73 researchers. The volume that calls itself an “Introduction” in fact represents the state-of-the-art in the field. It is only appropriate that the Slavonic manuscript tradition also finds a place in this context. Several sections of the book are dedicated to “Slavonic manuscripts” (Cleminson, 2015a, 2015b, 2015c; Sels, 2015;² Ambrosiani, 2015). From time to time, the authors are referring to the Glagolitic tradition specifically, but there is no explicit chapter to present or summarize it – the first indication that a lot of work is still required in this research area. This also corresponds with Per Ambrosiani’s conclusion:

The most important task seems to be the development of stable methods that can combine the best of the scholarly tradition with new technical possibilities, not only by producing new catalogues and descriptions of Slavic manuscripts, but also by integrating both earlier and more recent scholarship and making it widely available to both the multinational Slavic research community and international research outside the Slavic field (Ambrosiani, 2015).

1 Cf. *Summer school in the study of old books* (2010).

2 A case study on “Textual criticism and text editing”.

Two recent publications on Glagolitic manuscripts

A few weeks ago, I received the volume with a selection of papers presented at the 2016 Zadar University School *Authority, Provenance, Authenticity, Evidence* (2018), another appealing publication of Zadar University under the auspices of Mirna Willer, Anne J. Gilliland and Marijana Tomić. Two articles out of fifteen therein are dealing with the Glagolitic heritage: Milan Mihaljević (2018), one of the doyens of Glagolitic linguistics, comprehensively describes newly discovered manuscript fragments written in Glagolitsa dating from the 13th, 14th and 15th centuries, highlighting criteria for dating and identification. The other article, elaborated by Gordana Franov-Živković (2018), a great expert in the Glagolitic heritage of the Zadar region, introduces the interesting genre of “Confraternity Books”, a category of texts, which emerged in the Baroque period as a source for historiography. At the end of her article, the author presents an impressive phalanx of autochthonous manuscript sources of this kind in Glagolitic script. These two articles are of highest scientific quality, undoubtedly Glagolitic studies at its best.

As I immersed myself in these papers, some ideas came into my mind, since I noticed there is no perspective developed towards *Digital Humanities*. The question arose, whether it would make sense and eventually be of some advantage to migrate all these scientific results to the at times puzzling world of digital data or not. In my view, this is a strategic question of considerable weight. I have personally been confronted with this fundamental question and its problems for years, having to deal not only with Latin texts, but with such strange alphabets like Armenian or Syriac – which are presenting us with similar challenges of working on manuscripts written in less frequent alphabets. Thus, the questions are: What are the advantages of going digital? What is the price of such a decision? Where are the difficulties? Finally, who should carry out the work?

From shelf to screen – a multi-stage undertaking

The Zadar Pisana Baština-project

In 2015, the *Pisana Baština*-project to promote digitization of the written heritage of the Zadar region was developed by Marijana Tomić³ assisted by a

3 Marijana Tomić is a scholar tutored and widely coined by Mirna Willer.

handful of colleagues. The project was eventually launched on a practical level early in 2016 and has been running successfully since that time. Right from the start, it has become evident that digitizing manuscripts (or any document) goes far beyond working with a camera and taking photos.

Of course, digitized manuscripts are of huge advantage for researchers. Digital objects must be made accessible at least to the scientific community.⁴ The project had to harvest digital images, to allocate them unique and identifiable names and present them on a web portal. By the way, naming pictures – of sometimes unpaginated books and documents – is the very first step of metadata creation. The necessity of creating a web infrastructure and of metadata implied that the project had to seek the partnership of an IT expert.

However, representing single objects and presenting them online were by no means the only requirements.

The project started working with manuscripts in the collection of the Archdiocesan Library of Zadar. We were in the happy position of having a descriptive catalogue available to us in printed form. The catalogue can be best described as Pavao Kero's scholarly lifework (2015).⁵ We made the selection of manuscripts to be digitized with the help of this catalogue. A plenitude of information – on codicology, provenance, history, and contents – was registered there, which was eminently helpful for us. But the manuscripts themselves were not yet catalogued and searchable in a database. This was the moment where the dream of an electronic catalogue emerged. Alas! All data for this kind of catalogue would have to be extracted again, a task that could be only completed by hand.⁶

The *Pisana Baština* project is well underway and already generates excellent results as one can see on the website indicated above. However, there is one substantial item that has not yet been approached: encoding the full text of one of these manuscripts.

The Graz Summer School Hidden Libraries II

In July of 2018, we organized a summer school within the Utrecht Network of

4 It is an additional but not that easy decision on what kind of open access policy such an undertaking would like to follow, especially in case the project is funded by public means.

5 We were extremely saddened by the news that Don Pavao Kero, the local champion of Glagolitic studies, passed away in December 2018. We will keep in mind his open-mindedness and readiness for cooperation.

6 Marijana Tomić immediately started this work for all the manuscripts preselected for digitization according to the internationally valid cataloguing rules.

Universities under the title *Hidden Libraries II*. The central object of study was a collection of nearly one thousand parchment fragments, stored in a wooden box in perfect disorder and fully unclassified – remnants of a few medieval Latin books.⁷ It was one of the core aims of that summer school to unfold a way of how to proceed in order to get some of these pieces from “shelf [sc. box] to screen”. In a long-term perspective, we would like to virtually reconstitute the original book despite all its lacunae.⁸

Luckily, we found competent and high professional support through our partners from the *Austrian Centre for Digital Humanities*⁹ here in Graz: there was a codicological description of single fragments, there was the transcription of contents, there was the palaeographical assessment, etc. – all this done according to internationally established parameters. Then came the questions of what to do with the harvested data, of how and in which form(ats) to migrate them onto computer operable surfaces.

A complex and multifaceted enterprise

Dealing with manuscripts appears to be a complex matter. There is a whole chain of measures to be carried out from shelf to screen, starting with the retrieval of a manuscript (or any other document), the assessment of its conservatory state, the digitization, the creation of metadata, the ..., until it finally pops up on the surface of a computer or any other electronic device.

The following paragraphs try to provide a brief, simplified outline of what we thought would be advantageous for the processing of data in view of manuscripts with non-Latin alphabets, while remaining aware that this is only one way among a variety of possibilities. Having in mind the situation as laid out by the two articles quoted under item 2 (“Two recent publications”) and the experiences with the projects alluded to in the previous lines of this chapter, I would like to develop the idea of processing data on Glagolitic manuscripts so as to promote the motto “Glagolitics go online”.¹⁰ I will concentrate on three aspects: catalogue, text, and images.

7 A first autopsy made evident that at least 120 fragments belong to one single codex.

8 We are thinking of launching a research proposal to realize this plan.

9 *Austrian Centre for Digital Humanities* by courtesy of Georg Vogeler, Helmut Klug, and Roman Bleier. Helmut Klug made use of the material elaborated for and within the summer school in a series of lectures on management of digital projects during the winter term 2018/19, thus further advancing the idea.

10 In the last years we gained some experience with a similar task encoding Syriac texts for a searchable database, *Syriac-anaphoras* (Ephrem Aboud Ishac, Thomas Klampfl, Erich Renhart).

Digital Humanities

Such an ambition places the subject amid of what is summarized under the term of *Digital Humanities*. Handling data in the areas of electronic cataloguing, text editing and the use of digital images, we could apply self-invented and perfectly tailored proprietary IT solutions. However, these are necessarily of very limited value when it comes to interoperability. For this and other reasons, we are inclined to use standardized formats, applications, programmes and systems. Whatever solution we choose to best match our purposes, there is none without disadvantages. Thus, we always have to weigh the alternatives to find out what fits in best. This too, is a matter for occasional discussion.

Cataloguing

In practice, we are confronted with a wide range of catalogues forms, from poor inventory lists to most elaborated object descriptions as in Pavao Kero's catalogue quoted above. All of them are useful so as not to be urged to start from scratch. But even the most sophisticated printed catalogue has its limits when you want to make your books visible on the web or to virtually integrate your local library of Glagolitic manuscripts (or others) into the wide landscape of collections on the national or international level. I am vehemently favouring online catalogues – from which we normally can produce print versions.

The appropriate form of such catalogues would be the database. In choosing this path, we would have to give our information a standardized structure, which is readable for machines. Such databases are organized in different ways: as *relational databases* with firm and stable predefined fields or as XML-databases¹¹ which offer more flexibility. Both have their advantages and their disadvantages.¹² I clearly vote for XML-databases as they allow more flexibility in structuring units and in consequence can be better tailored to match the peculiarities and necessities of recording e.g. composite manuscripts. Additionally, “there is no limit to the amount of data to be stored within a given field” (Gippert, 2015b, p. 24).

Machine readable formats for cataloguing require standardized encoding of bibliographic data. Mirna Willer is a veritable champion in developing such standards worldwide. As a member and chair of the International Federation of

11 “XML is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable” (*XML*).

12 The two systems are concisely introduced and visualized in the fundamental article of Gippert (2015b).

Library Associations and Institutions (IFLA) Permanent UNIMARC Committee (IFLA, *UNIMARC Format*) she published extensively in the field of bibliographic standards, of records to be expressed in an XML syntax.¹³

Whatever the system of our choice, there are unsurpassable advantages to recording the Glagolitic library in a database. Gippert (2015a) summarizes:

- the data they contain can be retrieved dynamically [...],
- they can be linked to any other kind of electronic content in the web such as, for example, other catalogues, image repositories, or bibliographical materials [...],
- they can be steadily corrected, updated, extended and expanded.

Text editing

What we are able to do in terms of bibliographic information could also be applied to the *texts* contained in a manuscript or other kind of written document: try to make texts machine readable and edit them online or for print. Again, we would have to encode our texts corresponding to well defined standards. There is one indispensable prerequisite to do so, the use of an appropriate font. “Appropriate” means a Unicode-font, whose characters are encoded correlative with the universal standard (*UTF-8*). Fortunately, the Unicode block for the Glagolitic script is well developed (*Glagolitic (Unicode block)*).¹⁴

As mentioned before, we would prefer to use a markup language (XML) since this offers incredible possibilities of text processing. The most promising perspective for future presentation of texts in a digital ambiance seems to be expressed by the *Text Encoding Initiative (TEI)*, “a consortium which collectively develops and maintains a standard for the representation of texts in digital form. Its chief deliverable is a set of Guidelines which specify encoding methods for machine-readable texts, chiefly in the humanities, social sciences and linguistics.” TEI is an application of XML.

One of the sometimes tiresome tasks working with markup languages is the definition of *everything*: you have to tell the machine to do this for every single word. You would have to describe what you see facing a manuscript page, all units (lines, paragraphs, headers, etc.), all character properties You may even give additional information to every word (lexical, grammatical or any other).¹⁵ This process is called “tagging”, an essential step when encoding texts. However,

13 Actually, cataloguing the Glagolitic manuscripts of the Zadar Archdiocesane Library has already started according to these standards by M. Tomić, as mentioned above.

14 For the question with the representation of superscript and overbar signs in general cf. Gippert (2015b, p.21).

15 Cf. Gippert (2015b, 23) giving some samples of Georgian words.

we are not dealing here with boring work that we may consider to be redundant if compared with “simple typesetting”. This is the richness of a method as well, which allows us to create a more and more comprehensive lexicon of a language expressed in a certain alphabet. Having elaborated a good quantity of texts and the corresponding vocabulary in this way, such linguistic and grammatical data could be displayed automatically (with the help of so called “parsers”) when entering a word which is already available from earlier work.

We should possibly add here, that texts input according to the TEI guidelines can be exported to the web language HTML, and with the help of style sheets used to build plain text files, PDFs and so on. “The essential prerequisite is that the initial transcription should be accurate, and the primary encoding as comprehensive as possible in its informational content” (Cleminson, 2015, p. 83).

Handling Images

The question of digital imaging is dealt with in the *COMSt Handbook* in an introductory form (Gippert, 2015b, pp. 25-27). Of course, the techniques of taking photos from all parts of a manuscript and the awareness of resulting problems increased during the last 15 years. There are some technical standards which should be considered in order to obtain good results.

Our question here is primarily the operating of digital images and their connection to metadata. Once we have got the textual data in their encoded form, we may link them to images or to predefined areas of images. There are plenty of possibilities to connect metadata and images. Everything depends on what we want to have or can afford.

Considering the emergence of Glagolitic fragments (as mentioned above), we may ask whether or not it would be useful to create a central repository of images or to make the images stored in local memories accessible (at least) for researchers. This would no doubt immediately lead to the questions of quality control, copyright, and interoperability in general terms. We cannot provide answers to all eventualities here. But we would like to point out the *International Image Interoperability Framework* (IIIF), which “defines several application programming interfaces that provide a standardized method of describing and delivering images over the web, as well as ‘presentation based metadata’ (that is, structural metadata) about sequences of images.” This is to say, standards can be gained even with regard to the images we use, need or want to put for disposition for the scientific community – a reflection worth considering in a future perspective.

Towards a new horizon

The ideas expressed in this impromptu-like essay may cause some headache, since one might ask who can do all that? It is true; everybody dealing with ancient manuscripts has to be an expert in a range of areas: philology, history, codicology, liturgy, digitization and the like. There is no doubt that working in a digital environment would inevitably increase the workload. But there are some fruits we could gain which otherwise would not be reached: a national and/or international catalogue of Glagolitic manuscripts and fragments, a searchable database of all Glagolitic texts, an international bibliography on the subject, finally a gradually enriching lexicon – as a most welcome tool for ourselves and for other researchers.

The work can only be done by linguists and philologists who have a kind of affection towards the *Digital Humanities*. What has been achieved in this field with regard to other ancient text traditions is recorded by philologists in the *Comparative Oriental Manuscript Studies Handbook*, which I have quoted above several times. However, they are supported and assisted by very dedicated and sometimes passionate IT experts and cataloguers. Finally, we have to state, it is not that much extra work, even if we consider the tagging of every single word for an edition, since this is what we have to do anyhow when we translate and try to understand.

I am dreaming of a (possibly common) project to do some first steps, of a training to get an initial experience with the questions and the practical workflow. An excellent exercise to start with could be the cataloguing, textual analysis, and presentation of a Glagolitic fragment. This is not too voluminous, so one could yield results within reasonable time. The Department for Information Sciences of Zadar University, Mirna Willer's previous workplace, could be a privileged place to carry out such a project.

To conclude, let me quote Jost Gippert (2015b, p. 24) who phrased it in that way:

The more scholars show interest in these kinds of tools and methods, the more it is likely that we shall have them at hand for usage in the foreseeable future.

REFERENCES

- AMBROSIANI, P. (2015). Catalogues of Slavonic manuscripts. In A. Bausi et al. (eds) *COMSt Handbook* (pp. 500-502). Hamburg: Tredition.
- Austrian Centre for Digital Humanities, <https://informationsmodellierung.uni-graz.at/en/> (29-03-2019).
- Authority, Provenance, Authenticity, Evidence: Selected Papers from the Conference and School Authority, Provenance, Authenticity, Evidence.* (2018). M. Willer, A.J. Gilliland and M. Tomić (eds). Zadar: Zadar University.
- COMSt Handbook: Comparative Oriental Manuscript Studies: An Introduction.* (2015). A. Bausi et al. (eds). Hamburg: Tredition. [Renhart, E. (2017). Review. *Orientalistische Literaturzeitung*, 112 (1), 93-99].
- FRANOV-ŽIVKOVIĆ, G. (2018). Glagolitic confraternities' books of the Zadar area and their value as historical documents. In M. Willer, A.J. Gilliland and M. Tomić (eds) *Authority, Provenance, Authenticity, Evidence* (pp. 377-408). Zadar: Zadar University.
- GIPPERT, J. (2015a). Catalogues and cataloguing of oriental manuscripts in the digital age. In A. Bausi et al. (eds) *COMSt Handbook* (pp. 531-537). Hamburg: Tredition.
- GIPPERT, J. (2015b). Digital approaches to oriental manuscript studies. In A. Bausi et al. (eds) *COMSt Handbook* (pp. 12-27). Hamburg: Tredition.
- CLEMINSON, R.M. (2015). Encoding Text and Encoding Texts: Some Reflection on Theory and Practice. *COMSt Bulletin*, 1-2, 77-83.
- CLEMINSON, R.M. (2015a). Slavonic codicology. In A. Bausi et al. (eds) *COMSt Handbook* (pp. 235-251). Hamburg: Tredition.
- CLEMINSON, R.M. (2015b). Slavonic manuscripts. In A. Bausi et al. (eds) *COMSt Handbook* (pp. 55-57). Hamburg: Tredition.
- CLEMINSON, R.M. (2015c). Slavonic palaeography. In A. Bausi et al. (eds) *COMSt Handbook* (pp. 310-315). Hamburg: Tredition.
- Glagolitic (Unicode block)*, [https://en.wikipedia.org/wiki/Glagolitic_\(Unicode_block\)](https://en.wikipedia.org/wiki/Glagolitic_(Unicode_block)) (29-03-2019)
- Hidden Libraries II* (2nd-13th July 2018), <https://vestigia.uni-graz.at/de/lehre/summer-schools/2018-hidden-libraries-ii/> (29-03-2019).
- International Federation of Library Associations and Institutions (IFLA), *Universal Machine Readable Cataloguing (UNIMARC) format*, <https://www.ifla.org/unimarc> (29-03-2019).
- International Image Interoperability Framework (IIIF)*, <https://en.wikipedia.org/>

- wiki/International_Image_Interoperability_Framework#Image_API (29-03-2019).
- KERO, P. (2015). *Popis glagolskih kodeksa zadarske nadbiskupije* (2. prerađeno izd.). Zadar: Sveučilište u Zadru.
- MIHALJEVIĆ, M. (2018). Description of newly discovered Glagolitic manuscripts. In M. Willer, A.J. Gilliland and M. Tomić (eds) *Authority, Provenance, Authenticity, Evidence* (pp. 71-85). Zadar: Zadar University.
- Pisana Baština*, <https://pisanabastina.unizd.hr/> (29-03-2019).
- SELS, L.(2015). The Vidin Miscellany: translated hagiography in Slavonic. In A. Bausi et al. (eds) *COMSt Handbook* (pp. 448-452). Hamburg: Tredition.
- Summer school in the study of old books:proceedings*. (2010). M. Willer and M. Tomić (eds). Zadar: Zadar University.
- Syriac-anaphoras*, <https://syriac-anaphoras.org/> (29-03-2019) (Ephrem Aboud Ishac, Thomas Klampfl, Erich Renhart).
- TEI: Text Encoding Initiative*, <https://tei-c.org/> (29-03-2019).
- UTF-8*, <https://en.wikipedia.org/wiki/UTF-8> (29-03-2019).
- XML* (eXtensible Markup Language), <https://en.wikipedia.org/wiki/XML> (29-03-2019).

IZDANJA GLAGOLJSKIH TEKSTOVA I PERSPEKTIVA DIGITALNE HUMANISTIKE

KLJUČNE RIJEČI:

*digitalna humanistika, Text
Encoding Initiative, XML, glagoljski
rukopisi*

SAŽETAK

*Glagoljska baština još nije dostupna online ili je dostupna vrlo ograničeno. Ovaj članak pokušava formulirati perspektivu premošćivanja jaza uz poticanje uporabe jezika za označivanje (XML) za katalogizaciju i tekstualna izdanja. U više se navrata poziva na priručnik *Comparative Oriental Manuscript Studies Handbook*, 2015. koji se bavi susjednim drevnim pisanim tradicijama Sredozemlja i njihovim postignućima u digitalnoj humanistici.*