

**Tomaz Erjavec**

Dept. of Knowledge Technologies,  
Jožef Stefan Institute, Ljubljana, Slovenia

**Matija Ogrin**

Institute of Slovenian Literature and Literary Studies,  
ZRC SAZU, Ljubljana, Slovenia

## THE eZISS DIGITAL LIBRARY OF SCHOLARLY DIGITAL EDITIONS OF SLOVENIAN LITERATURE

**Abstract:** The paper presents the eZISS digital library (<http://nl.ijs.si/e-zrc/>), which currently contains eight critical editions of Slovenian literature, all original works in scholarly editing. Some of the editions represent the most important milestones in the history of the Slovenian language, such as the Freising Manuscripts, the earliest preserved writings in Slovenian and the Škofja Loka Passion Play, the oldest play written in Slovenian. The editions present the text by means of several methods and perspectives, such as digital facsimile, diplomatic transcription, edited text, in some cases translations, glossaries and the like. One of the methodological goals was to distinguish clearly, on the one side, the documentary or diplomatic view of the text from the critical or edited view on the other side. The paper presents the methodology used in the making of these editions, esp. the connection of these methods with the technology of text markup. The encoding of the editions is based on the Guidelines for Electronic Text Encoding and Interchange (TEI), where the technology standards are connected with the classical philological methods of scholarly editing. This approach has proved successful for publishing of the eZISS editions and has led to an increased interest for research and publishing of primary sources of the Slovenian literature.

**Keywords:** Digital scholarly editions, Slovenian literature, Slovenian language, Text Encoding Initiative

### 1. Introduction

Scholarly editions present literary texts to the reader by means of several philological methods, which is why these editions establish the text as the object of literary study. In this sense, scholarly editions in Slovenian studies – as in every national literary history, help create the foundation of literary scholarship as they, and to a large extent constitute its subject matter. The preparation of a complex scholarly edition must take into account the details of the primary text, the biography of its author and the socio-cultural context of the time when the text was created, thus drawing upon the widest possible influences on the text being studied. A text-critical edition thus implies first and foremost thorough research of a work or a complete opus, strictly observing philological methods until its final publication. When a scholarly edition discovers, (re)constitutes and offers the primary text as an object of literary studies, it also determines, to a certain extent, the evolution, authenticity and timeliness of the studies themselves.

Nevertheless, the history of literary studies in Slovenia shows that endeavours in scholarly editing gained impetus only as late as 1946, when the series of critical editions *The Collected Works of Slovenian poets and writers* [10] started. Another outstanding achievement in the field was the more recent text-critical edition of the *Freising Manuscripts* [1] which was used as the basis for our digital edition of the *Freising Manuscripts* [6].

The eZISS project started in 2001 [2], with the goal of giving a new impetus to text-critical

work by linking the classical philological tradition in textual criticism in Slovenia with IT methods for text-processing and presenting the results of the research. In the years since, eZISS has published eight editions, which is a small number for today's digital libraries – but, as mentioned, these are not simply digitised works, but precisely transcribed, critically (semantically) encoded and commented editions, authored in the digital medium. The developed methodology of producing digital editions proved to be fruitful also for other similar philological and historical disciplines. This is why it has also been adopted in the wider Slovenian context, e.g. for the *Unknown Slovenian Manuscript descriptions* (NRSS, 2011), *Digital scholarly publications* (eZMono, 2011), the on-line *Slovenian Biographical Lexicon* (SBL, 2009) and the AHLib digital library and corpus of historical Slovene books (AHLib, 2012).

## 2. The structure of a scholarly edition

In the digital world, the base object of a scholarly edition is captured in the collection of the images of the written or printed pages, which is then typically described at least in three ways. The first is the meta-data, i.e. bibliographic and other descriptions of the pages that constitute it. Especially for manuscripts of great value, these descriptions can be highly detailed, giving information about the repository, provenance, history, physical conditions, collation, etc. of the manuscript. A text-critical edition will typically also contain commentary on the text, in the form of notes and separate chapters of the edition, together with some front-matter, bibliography and appendices. Furthermore, the vocabulary of the text can be analysed, and a glossary of the old words or word forms included in the edition, possibly using a dictionary structure.

The other two “descriptions” of the object are the diplomatic and critical transcriptions, which interpret in a textual form the object or its image. The diplomatic transcription tries to reproduce the text, i.e. captures, as much as possible of the original, including the errors, corrections and later additions, damage and different hands that have written on a manuscript, printer's marks in books, etc. The production of a diplomatic transcription forces the encoder to analyse all the dubious parts of the text, where it might be illegible, overwritten by later hands, crossed-out, etc. With such an analytic approach the diplomatic transcription follows, at least to an extent, the genesis of the text, and makes this genesis explicit. For this reason, it is an essential method in the field of historical scholarly editing [8].

It is not the purpose of the diplomatic transcription to decide which parts of the text are more or less relevant, correct or better. Such decisions are made in the critical transcription, which interprets the text by performing some sort of normalisation on it: it will take the authors' emendations into account, choose the best “path” through the text, correct and encode errors, expand abbreviations, modernise the punctuation, spelling, etc. The critical transcription is oriented toward the reader, by making the literary work better understandable. As there are different types of readers, there are also different critical transcriptions, possibly several per text, e.g. in using contemporary alphabet or spelling, giving the (reconstructed) phonetic transcription, or even translating the text. Another type of text-critical analysis can be performed by juxtaposing two or more sources, say two editions of the same text, and noting and analysing the differences via the so called text-critical apparatus. A text-critical edition might also have associated further materials that shed light on the text, e.g. the audio recording of the read or performed text.

As can be seen, a text-critical edition has a complex structure made of possibly many parts (facsimile, meta-data, diplomatic transcription, critical transcriptions), and these, in order to relate

them, must be interconnected: it must be obvious which part of a transcription describes which page or page area, which two lines of different transcriptions are describing the same text, and which read (audio) segment corresponds to which sentence in the text.

### 3. The eZISS editions

From the very first edition in the mid 2004 until the latest in 2011, eZISS saw the publication of eight scholarly editions. Each of them was conceived and prepared according to historical and textual peculiarities of the given text so that the concept of the edition would be suited to the nature of the text as much as possible. For us, „as much as possible“ means that the edition should enable all the views or perspectives on the text that are relevant for its reading and study – from the facsimile to several options of normalisation and even translation. The texts themselves range from the 10th century Freising Manuscripts up to the 20<sup>th</sup> century novel by Izidor Cankar and avant-garde poetry by Anton Podbevšek. This variety of texts implies the variety of editorial concepts, which is outlined in Table 1. The table gives, for each edition, its components and mark-up features, further elaborated in the next section.

**Table 1:** Overview of the eZISS editions

Edition	Components	Mark-up features <sup>1</sup>
Freising Manuscripts, 2007	facsimile plus <i>editio variorum</i> (several diplomatic, critical and phonetic transcriptions by earlier and modern Slavonic scholars in full text), translations into five modern languages; dictionary; audio recording, glossary, commentaries; several views for parallel presentations of altogether 17 versions of the Manuscripts.	special encoding for medieval characters; parallel views; sentences and segments aligned with audio recording
Škofja Loka passion play, 2009	facsimile, diplomatic transcription, critical transcription; commentaries, video recording of the performed play.	encoding of ms. hands, video recording synchronised with transcriptions
Municipal swearing-in texts, 2009 (revised 2011)	facsimile, diplomatic transcription, critical transcription; historical commentaries, vocabulary analysis.	
Swearing-in texts of borough and professional communities, 2011	facsimile, diplomatic transcription, critical transcription; historical commentaries, vocabulary analysis.	encoding of ms. hands
Correspondence of Žiga Zois, 2004 (revised 2007)	facsimile, diplomatic transcription, translation in Slovenian; commentaries.	
Slomšek: Three sermons on language, 2004, (revised 2007)	facsimile, diplomatic transcription, critical transcription; commentaries	
Izidor Cankar: From the road, 2007	two editions of the novel (1913 and 1919), presented separately as well as in a parallel view with explicit variants; commentaries	variants encoded as with text-critical apparatus and related elements

<sup>1</sup>Only special mark-up, other than regularly used tags for text-critical structures, is mentioned here.

Anton Podbevšek: Collected poems, 2007 (revised 2010)	facsimile of several primary sources (including manuscripts and the author's own copy of his book with substantive corrections), transcriptions; parallel presentation of versions; commentaries	encoding of white-space (visual poetry)
---	--	---

Each edition is the result of fundamental research on primary sources, related historical context, persons and additional material. Projects in scholarly editing in Western countries are, typically, focused on a particular epoch or even on a particular writer or on some specific topic, corpus separatum etc. The number of editions and the extent of the published materials are typically larger, but on the other hand, the historical and text-critical issues per project, are consequently, similar for entire corpus of texts. With the eZISS editions, the situation is rather the opposite. The editors had to cope with extremely diverse types of editorial problems: the medieval *Monumenta* touch upon quite different text-critical issues that does the 18th century baroque *Passion play*. The Slovenian 17th century *Oaths* were edited and presented in a considerably different way than the early-19th century *Sermons* by Anton Martin Slomšek, etc. The texts in these editions also span a number of literary genres, e.g. rhetorical prose (sermons), a dramatic passion play, a novel, modern poetry, documentary materials, cosmopolitan epistolary collection, etc. The types of editions, needed to encompass these materials, include an expanded editio variorum (*Freising Manuscripts*); elements of genetic editing are to be found in the edition of the *Passion play*, and in Podbevšek's poems; the Municipal swearing-in texts and the *Zois* correspondence represent different kinds of classical documentary editions, etc. Each text came up with singular problems and each eZISS edition tried to be an answer to it.

#### 4. Encoding of the editions

Such a multitude of views on the text demands a plethora of methods and approaches on the part of the text-encoding. For eZISS we adopted an already developed set of encoding practices, which has been, furthermore, made for describing texts for scholarly purposes, for the simplest to the most complex. The Guidelines that we used were developed by Text Encoding Initiative [9], which has managed to transfer, incorporate and apply a large part of the philological tradition and methods to the digital medium. TEI can be described as:

- a freely available set of guidelines for humanities-oriented text encoding, which uses XML (Extensible Markup Language) as its base standard;
- an international consortium with the purpose of the maintenance and development of the Guidelines;
- an international group of projects that are using the TEI Guidelines.

The TEI Guidelines for Text Encoding and Interchange define and document a markup language for representing conceptual structures of texts. They are couched as a modular and extensible XML schema with extensive accompanying documentation, and are available under an open licence. The first public version of the TEI Guidelines (so called TEI P3) was released in 1994, and has been, to correct mistakes and keep up with technological development, updated twice: TEI P4 was released in 2002, and TEI P5 in 2007 ([TEI, 2007](#)). In eZISS the older editions use the TEI P4 version, while the last three use TEI P5.

The TEI Guidelines define a few hundred XML elements and attributes for marking-up a wide variety of texts. The definitions of the elements and attributes are expressed in modules, each of which

covers a certain text type or interpretative apparatus, and each module is described in detail in a dedicated chapter of the Guidelines. To arrive at an XML schema, which is needed to formally validate a TEI document, these modules can be combined, as well as extended.

For the eZISS library we use XML schemas which – depending on the specific edition – use a rich collection of TEI modules. We list below some of the more interesting ones:

- *core*: common data  
This module is required for any TEI-conformant document – it defines generally useful elements such as paragraph, page break, highlighted text, bibliographical data, but also elements for a basic description of other text types, e.g. speaker, speech, and stage direction for drama.
- *header*: common meta-data  
This module is also required for every TEI document – it defines the structure of the TEI header, i.e. the element that holds the meta-data, such as its title, author, source description, editorial policies, use of elements, revision description, etc.
- *textstructure*: text structure  
This is another, and the last, required module and defines elements that describe the structure of a document, such as front matter, divisions, appendices, etc.
- *msdescription*: manuscript descriptions  
The module defines elements for a detailed description of manuscripts, e.g. details of the repository, description of their content, physical description (material, foliation, collation, history of the manuscript), a description of the hands, etc.
- *transcr*: transcription of primary sources  
This module defines elements that enable the inclusion of facsimiles into the documents, as well as elements used in diplomatic and critical transcriptions, e.g. deleted or added text, abbreviations and expansions, damaged and unclear passages, etc.
- *textcrit*: textual criticism  
The module gives elements for the text-critical apparatus, which enables the description of the differences between individual editions of a particular work.
- *linking*: linking, segmentation and alignment  
This module defines, inter alia, attributes for linking of elements; in the eZISS editions we use it primarily to connect together (lines, paragraphs) of the various transcriptions.
- *dictionaries*: dictionaries  
The module enables the encoding of dictionary structures, such as the dictionary entry, grammatical properties, the definition, translation, examples etc.
- *gaiji*: special characters  
The module enables the definition and use of special characters not present in Unicode. Because of the universality of the Unicode character set, cases where we need this module are fairly rare. Still, in the diplomatic transcription of medieval manuscripts we do find non-standard characters which can be, with the help of the *gaiji* module, rendered exactly with a special font which uses the Unicode Private Use Area, or as standard Unicode approximations with regular fonts.

In addition to using the TEI Guidelines we also used other encoding standards in the digital source of the eZISS editions. In particular, two editions also contain multimedia elements. The *Freising Manuscripts* have an integrated sound recording of the reading of the manuscript according to the phonetic transcription. Here the speech signal was segmented by sentence and clause, and the link to the appropriate speech segment simply marked on the individual sentence or clause element. The *Škofja Loka Passion Play* has an associated video performance, which is also available synchronised

with the diplomatic or critical transcription, realised in SMIL (Synchronized Multimedia Integration Language).

### 5. Access to the editions

If the documents encoded in TEI / XML represent the canonical form of the digital edition, which should be formally correct, interchangeable between applications and platforms, and resistant to technological change, the format itself is not suited for actual reading of the editions. For this purpose, HTML is much better, as this is the encoding designed for presenting text on the Web.

The conversion from TEI to HTML was made separately for each edition via scripts written in XSLT, which is a W3C standard, just as XML is. For the older editions, encoded in TEI P4, we wrote a dedicated XSLT script for each edition, to best display its specifics. For editions in TEI P5 we took a different route. With a relatively simple XSLT script each edition is converted to a simplified and representation-oriented TEI, while the actual conversion to HTML was performed by standard TEI XSLT scripts which are available on the TEI Consortium Web pages.

In addition to the HTML rendering of the texts of the editions themselves, we also converted to HTML the TEI headers of each individual edition. As mentioned, the TEI header contains the metadata of the edition, i.e. it documents the edition and uses a large number of specialised TEI elements for this purpose. In addition to classical bibliographic data we find here also the list of all TEI elements used in the text with their frequency, the description of special characters, detailed manuscript descriptions, etc. The TEI headers are, again with an XSLT script, converted to HTML. Additionally, the TEI elements used are localised to Slovenian; to this end, we wrote a TEI document which gives, for each TEI element a Slovenian gloss, and the conversion uses this localisation information [4]. The TEI headers themselves also give all the free text both in Slovene and in English, so it is possible to make them available in HTML in two versions – for Slovenians in their native language, and for the others in English.

If we compare the eZISS editions to similar efforts abroad, which also use TEI as their base encoding, we see that the main difference is in the presentation. While eZISS uses static HTML files, most others will present the editions in the scope of a content management system, which allows for greater versatility in searching and presenting the editions. However, static HTML does have the advantage of requiring minimal software on the server side, as well as supplying persistent identifiers for particular editions and their constituent parts.

The top-level HTML page of the eZISS library is also available both in Slovenian and English, and contains the introduction to the project, the bibliography and pointers to the individual editions. Each individual edition is then composed of the following parts:

- the home page with a short introduction to the edition, both in Slovenian and English, and links to the edition;
- the TEI header in HTML with descriptive names of each used TEI element in Slovenian and English;
- the edition in static HTML, for browsing and reading in Web browsers;
- the canonical edition in XML / TEI, together with the XML schema and XSLT scripts for conversion to HTML;
- compressed files in zip format, which enable downloading the edition (only the text, or in their entirety with associated facsimiles or multimedia content), so that the edition can be perused also without an internet connection.

The complete editions (TEI, HTML, XSLT) are all available under one of the Creative Commons – Attribution licences – this enables copying, processing and further distribution of the library, in most cases with the possibility of derived works and commercial use, as long as the authorship of the edition is acknowledged. Some editions also specify Share-Alike, meaning that, if the edition is modified, it should be then shared under the same conditions as the original, or No-Derivatives, meaning that the re-distribution of modified copies is not allowed.

The eZISS Web page also offers a service for converting up-loaded Word files to TEI [3]. Namely, it turned out that authors of text-critical editions are, as a rule, not very interested in investing the time to learn about XML and TEI, so it is easier for them to author the editions in Word, which is then automatically converted to basic TEI. From there on, this TEI can be semi-automatically converted to the final edition. The conversion supports typical structures such as divisions, headings, tables, footnotes, highlighted texts etc., but also more specialised TEI elements. This is achieved by supplying a special .dot file on the Web page, which offers a number of *tei:* prefixed styles. So, for example, if a series of paragraphs is marked with the style *tei:sp*, then the series will be converted into the TEI element `<sp>` (for “speech” in a play) and, inside the speech, the first paragraph will be marked up as `<speaker>`, with the others as `<l>` (line in a speech).

TEI is of course much more constrained than is Word, so it is easy to produce Word documents that do not result in valid TEI, or in a TEI that is different from what the author intended. The first problem is solved by automatic validation of the resulting document, which pin-points the problematic areas of the source Word. To alleviate the second problem, the Web service also supports the conversion of the TEI document to HTML, so that the user can visually compare the HTML with the source Word, and modify the latter in case of substantive differences.

The works for eZISS are well-indexed by Google, in part probably due to the fact that they are linked to from their corresponding articles in Wikipedia. We monitor the access to the pages of the eZISS digital library with Google Analytics, which shows that our editions are, on the average, daily visited by 50 users, half of these from Slovenia. About half of all the users find the digital library or one of its individual editions through Web search engines (where the most common search terms are »brižinski spomeniki« (Freising Manuscripts) and »anton martin slomšek«), and a third by following links from elsewhere. About a fourth of the users visit the library more than once, and about 6 % (in the last year about 1,000) stay on the site for more than 10 minutes.

## 6. Conclusions

The eZISS library, <http://nl.ijs.si/e-zrc/>, has started as an experiment in how to transfer classical philological methods in producing text-critical editions into a digitally-centred publishing environment. The experiment could succeed only through integration of three components: textual criticism, the tradition of Slovenian literary studies and the technologies related to the digital mark-up of texts. It turned out that this new method of preparing scholarly digital editions is a fruitful one, as it opens useful and modern paths to (old) philological goals. The XML / TEI technologies have shown great potential to provide not only support and tools for basic research in Slovenian literary studies, but also to give new impetus to them. The interest of the humanities scholars could thus, from the initial problems with technology, soon turn to the primary sources – old prints and manuscripts – with a newly awakened interest for their study.

We tried to bring the eZISS editions as close to the user as possible. This is why we adopted a precise and standardised encoding of the editions, and published them under very liberal licence

conditions, with which eZISS leads the way not only in Slovenia but also in the wider region. The eZISS editions use various methods of analysis and text transcription, as well as a combination of transcriptions and other media, so as to enable diverse views and approaches to the text. It was possible to do this without reinventing the wheel because we based our efforts on the work of others, esp. the widely adopted international standards and recommendations for text encoding – Unicode, XML, and TEI. This ensures not only the longevity and platform independence of the digital editions but also the methodological precision and, not the least, a positive contribution of the Slovenian literary component in globalised trends in the digital sharing of scientific data.

### References

- [1] Bernik, F., Grdina, I. et al. (ed.), 2004: *Brižinski spomeniki – Monumenta Frisingensia. Znanstvenokritična izdaja*. Ljubljana, Trst: Mladika, SAZU, ZRC SAZU.
- [2] Erjavec, T., Ogrin, M., Faganel, J., 2004: E-Slomšek: a TEI encoding of a critical edition of 19th century Slovenian rhetoric prose. *Pregled Nacionalnog centra za digitalizacijo*, vol. 5, 31-41.
- [3] Erjavec, T., 2007: Architecture for editing complex digital documents. In *Proceedings of the First International Conference “Digital information and heritage”*. Zagreb: Odsjek za informacijske znanosti, Filozofski fakultet, 105-114.
- [4] Erjavec, T., 2010: Text Encoding Initiative Guidelines and their Localisation. *Infoteka*, 11/1, 3-14.
- [5] David C. Greetham, 1995: *Scholarly Editing. A guide to Research*. New York: MLA.
- [6] Ogrin, M., et. al (ed.), 2007: *Brižinski spomeniki – Monumenta Frisingensia. Elektronska znanstvenokritična izdaja*. ZRC SAZU. <<http://nl.ijs.si/e-zrc/bs/>>
- [7] G. Thomas Tanselle, 1992: *A Rationale of Textual Criticism*. University of Pennsylvania Press.
- [8] G. Thomas Tanselle, 1995: The Varieties of Scholarly Editing. In: *Scholarly Editing. A guide to Research* (D. C. Greetham, ed.). New York: MLA, 9–32.
- [9] TEI Consortium, 2007: *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 2.1.0. Last updated on 17th June 2012. <<http://www.tei-c.org/Guidelines/P5/>>
- [10] Zbrana dela [series, A. Ocvirk et. al., ed.], 1946–2011: *Zbrana dela slovenskih pesnikov in pisateljev*. Ljubljana: DZS, Založba Obzorja, Littera, ZRC SAZU.

### Related projects

- AHLib, 2012: AHLib digital library and corpus: <http://nl.ijs.si/ahlib/>
- eZMono, 2011: Elektronske znanstvene monografije. Ljubljana: ZRC SAZU: <http://ezb.ijs.si/fedora/get/ezmono:ezmono/VIEW/>
- NRSS, 2011 (Ogrin, M., ed.): Neznani rokopisi slovenskega slovstva 17. in 18. stoletja. Ljubljana: ZRC SAZU: <http://ezb.ijs.si/nrss/>
- SBL, 2009 (Vide, P., ed.): Slovenski biografski leksikon 1925–1991. Elektronska izdaja: <http://nl.ijs.si/fedora/get/sbl:sbl/VIEW/>

[tomaz.erjavec@ijs.si](mailto:tomaz.erjavec@ijs.si)

[matija.ogrin@zrc-sazu.si](mailto:matija.ogrin@zrc-sazu.si)