Journal of the Text Encoding Initiative

Issue 4 | March 2013

Selected Papers from the 2011 TEI Conference

A Register of Early Modern Slovenian Manuscripts

Matija Ogrin, Jan Jona Javoršek and Tomaž Erjavec

Electronic version

URL: <http://journals.openedition.org/jtei/715>

DOI: 10.4000/jtei.715

ISSN: 2162-5603

Publisher

TEI Consortium

Electronic reference

Matija Ogrin, Jan Jona Javoršek and Tomaž Erjavec, « A Register of Early Modern Slovenian Manuscripts », *Journal of the Text Encoding Initiative* [Online], Issue 4 | March 2013, Online since 08 March 2013, connection on 08 May 2019. URL : <http://journals.openedition.org/jtei/715> ; DOI : 10.4000/jtei.715

This text was automatically generated on 8 May 2019.

TEI Consortium 2013 (Creative Commons Attribution-NoDerivs 3.0 Unported License)

A Register of Early Modern Slovenian Manuscripts

Matija Ogrin, Jan Jona Javoršek and Tomaž Erjavec

AUTHOR'S NOTE

We wish to thank the numerous monasteries, archives, libraries, and individuals, especially † Fr. Felicijan Pevec OFM, for making accessible the manuscripts in their keeping. The work reported in this paper was funded by the Slovenian Academy of Sciences and Arts and by the Slovenian Research Agency project "Neznani rokopisi slovenskega slovstva 17. in 18. stoletja."

1 Introduction

- 1 A scholar who is confronted with early modern manuscripts—and wants not only to read them but also to understand their genesis, to locate their origin, or to determine the sociocultural context of their author(s)—runs into several problems. Early modern manuscripts, just like medieval ones, hide their manifold messages. The complex historical and cultural phenomena materialized within them can only be observed and interpreted by means of precise analytical methods.
- 2 While the fundamental methodological framework of research on early modern manuscripts is similar to that of research on medieval manuscripts, there are important differences. Paleographic analysis of early modern manuscripts is limited because conventional traits of the script (which can be dated) are dominated by idiosyncratic traits of the author or the scribe (which cannot be dated). Hands become more and more individual, and when they cannot be identified, they offer little or no grounds for dating the manuscript. In modern manuscripts the precise layout of pages, well established as it was in the medieval tradition, became more loose and accidental, and often subject to

writer's arbitrariness. Still, the basic framework of codicological and paleographic methods is the same for medieval and modern manuscripts. Both kinds encompass "hidden" messages that, at first glance, are invisible: for instance, watermarks bear important temporal information in modern manuscripts, collation of quires is often highly relevant for textual aspects, and analysis of binding may contribute to the analysis of provenance.

- 3 The second important trait that the medieval and early modern manuscripts share, at least to some extent, is a similar notion of "publication" implied by their authors. In many European cultures and specifically in some parts of the national "literary systems," early modern manuscripts were regarded as publications (Love 1993, Moureau 1993): they were borrowed, read, and copied in a social environment which was much larger than the private environment of their original authors, and manuscripts actually had the nature and function of public texts. Even as late as the 18th century, deep in the era of print culture, manuscript culture not only survived but for several reasons also remained the prevailing medium of existence of (literary) texts. This holds true not only for typographically specific literary contexts such as Hebrew and Greek (McKitterick 2004, 11) but also for Irish (Ní Úrdail 2000), Icelandic (Driscoll 1997), and other "small" literatures, including Slovenian.
- 4 As in many national philologies, in Slovenian literary studies of the last 100 years a strong division between printed and manuscript books was taken for granted. This is a consequence of the commonly held view that the era of manuscript culture—and hence the importance of manuscripts—ended with the advent of print. As McKitterick makes clear, this misinterpretation was the reason why, in the minds of many modern scholars, "printing and manuscripts were divorced both in their organisation and in their study," and "as a consequence, modern bibliographical (and therefore historical and literary) study has been weakened" (McKitterick 2004, 22).
- 5 This also holds true for Slovenian literary history. Many Slovenian early modern manuscripts have never entered into the record of literary studies, and others, though of considerable relevance, have only been given a sketchy treatment. Still, manuscript codices, quires, and single leaves complement significantly the image of Slovenian baroque and enlightenment literature and culture in general. In the 17th and 18th centuries, many important modern textual genres appeared for the first time. Beside traditional medieval genres such as the theological treatise, sermon, and hymn, new types of texts appeared in the national language: passion plays (early modern drama), meditative prose, lyrical poems, liturgical songs, folk songs, collections of proverbs, and various texts from the apocryphal tradition. Some of these genres—folk songs, proverbs, and to some extent even church songs—had existed only in oral tradition. In early modern manuscripts, these genres find their way into writing for the first time—what can be called the rise of Slovenian popular literature. With this rise, not only did a variety of new textual types appear but the social base and the cultural context of the manuscripts' authors changed significantly as well. Beside ecclesiastics, secular writers and even many autodidacts and learned peasants produced astonishing "new" genres—with strong medieval elements, of course—such as prophecies, oracles, and spells.
- 6 The number of Slovenian-language manuscripts preserved from 17th and 18th centuries is not large. If, beside manuscript books, we also count as individual manuscripts small quires and partial leaves (such as charters, oaths, and letters), the total number may hardly exceed one thousand. If we only count substantial quires and books, the number

probably would not exceed four hundred. For historical reasons, most Slovenian authors wrote in German and Latin until the mid-19th century, so the few the surviving texts in Slovenian are valuable for literary, linguistic, and other studies.

- 7 These are the reasons why early modern manuscripts are a relevant object of research in humanities disciplines including Slovenian national history, language, literature, religion, and folklore. There is a need for a systematic, methodologically consistent scholarly account of all these materials. We want our research in manuscripts to be uniform, conceptually clear, machine-readable, reusable, and well documented. To record our own research and descriptions and to enable further analysis and presentation of these manuscripts, we developed a digital repository, the Register of Early Modern Slovenian Manuscripts, where the materials are available for searching, reading, and downloading¹ and all are uniformly encoded in TEI P5, for the most part using the TEI module for manuscript description.
- 8 As the extent of surviving manuscript material in Slovenian is relatively small, it is realistic to aim to describe all the surviving items. To date, the Register contains descriptions, bibliographical references, and digital facsimiles of one hundred manuscripts, plus a list of more than one hundred further manuscripts documented in different sources but not yet located (many of which are probably lost). The resulting Register is a small collection but nevertheless can serve as a starting point for research in the field.
- 9 This paper presents the Register from several perspectives, current as well potential. Section 2 presents the encoding of the materials, section 3 describes the online presentation and query system based on the Fedora Commons repository software, section 4 discusses the temporal and geographical aspects of the register in connection with their further formalization, and section 5 gives some conclusions and directions for further work.

2 The Encoding of the Register

- 10 The Register plays a dual role. Its first task is to give basic information on the existence of manuscripts, like a catalogue. The second goal is more ambitious and builds upon the achievements of the first one: we wish to record and make explicit the features that go beyond the manuscript as a material object, describing instead its textuality, authorship, and relationship to other manuscripts. Which textual genres are present in the manuscript? What was the context that gave rise to its genesis? Are there manuscripts written in the same hand in the register? Or are there manuscripts with the same text, but of different hands? Answering questions like these can require complex markup as provided by the TEI module for manuscript description. To sum up, the Register should:
 - give detailed and reliable manuscript descriptions, including the description of its content, materiality, and origin
 - encode and describe the identifiable manuscript hands in the collection, especially for the manuscripts of distinguished authors
 - explicitly encode and make searchable the genre(s) of each manuscript's text(s) as well as its sociocultural background
- 11 Using these criteria, we encoded descriptions of manuscripts from a number of libraries, private and public, in Slovenia and abroad, mostly in monasteries. Many of the

manuscripts, mentioned in earlier scholarly studies without clear citations to their respective repositories, could not be found, but we wished to collect and record all attainable information about them, apart from actual manuscript descriptions. For this reason, the web portal is composed of three parts:

1. The actual Register of manuscripts, which typically includes facsimiles and detailed manuscript descriptions and currently contains 100 manuscript descriptions and 7,011 pages of facsimiles.
 2. For manuscripts not yet found, an index of basic manuscript citations (collected together as "descriptions") with links to scholarly literature where they are mentioned or described—currently 176 descriptive citations.
 3. A list of bibliographical items where the manuscripts are mentioned—currently 220 units.
- 12 Additionally, the Register contains a detailed TEI header, giving the metadata for the complete Register; <front> matter documenting the aims, structure, encoding practices, and other aspects of the Register; and <back> matter containing the list of identified persons with links to their hands in the manuscript descriptions.

```

<msItemStruct class="#nrss_type_prose #nrss_type_sermon #nrss_type_varia
#nrss_aut_folk_carinth">
  <title xml:lang="slv" type="original">Branie od Antekrista inu od
niegoveh
    foush prerokou tude od konzhanja tega svieta Od gore uftajenja teh
mertleh
    inu od te Poshliednie Sodve Boshie Kai sa en vrtil zhries is vollane
inu Kai
    sa en vrtil zhries te ferdamane Prishu bode Preshribanu vtam letu
1780.</title>
  <title xml:lang="slv" type="editorial">Branje od Antikrista inu od
njegoveh
    fouš prerokov</title>
  <incipit xml:lang="slv">Krodsko popisane tega shiulena inu sapopadk
pridna
    poboshnosti Floriona posebnega pomozhnika uneuornosti tega ognja
predlia pa
    letania suoshena shnegouega shiulena.</incipit>
  <explicit xml:lang="slv">dusha noj telu bash Pogubu kna bash uzh u
shiulou
    Millosti, gnada bashja je sgublena dusha bo u selej pezhena kna bo
uezh
    osdraulena</explicit>
  <listBibl>
    <bibl copyOf="#bibl_123">
      <biblScope type="pp">448, 449</biblScope>
      <note type="bibl">Omemba in opozorilo v opombi, da je rkp morda
povezan
        z okolico oz. ozemljem gospostva Grafenstein.</note>
    </bibl>
  </listBibl>
  <textLang xml:lang="slv"/>
</msItemStruct>

```

Example 1. An example of `<msItemStruct>` element. The `@class` attribute has several values, pointing to the respective categories in the typology.

- 13 The encoding for each manuscript, the `<msDesc>`, is divided into blocks provided by the TEI Guidelines: the identifier of the manuscript (`<msIdentifier>`), its contents (`<msContents>`), its physical description (`<physDesc>`), and its history (`<history>`). Besides these, we used the `<additional>` element for administrative data about the record itself: who recorded or revised it and when. For analysis and description of the manuscript contents in `<msContents>`, we provide a `<summary>` and (a series of) detailed `<msItemStruct>` elements; an example is shown in example 1. The most salient encoding practices of the Register are the following:

- For the encoding of textual genres, we provide a typology of texts, encoded as a `<taxonomy>` with nested `<category>` elements in the TEI header, as exemplified in

example 2. Each manuscript description is then linked, via the @class attribute, to the appropriate categories of the typology.

- Similarly, we provide a typology of social contexts in which the manuscripts were written, with such authors as monastics (with further divisions into members of specific monastic orders), secular clergy, laypeople, civil offices, and peasants.
- The hand of every recognized author is encoded by means of a <handNote>, e.g. <handNote scribe="Rogerij">, so that a list of hands can be generated and combined with further information about an author or scribe.

```
<category xml:id="nrss_type_poem">
  <catDesc>
    <term xml:lang="slv">pesem</term>
    <term xml:lang="eng">poem</term>
  </catDesc>
</category>

<category xml:id="nrss_type_songbook">
  <catDesc>
    <term xml:lang="slv">pesmarica</term>
    <term xml:lang="eng">songbook</term>
  </catDesc>
</category>
```

Example 2. The typology of literary genres in the TEI header.

- 14 All these encoding practices are closely tied to analytical distinctions, decisions, and interpretations. In some cases, deciding what an autonomous <msPart> is, instead of a more regular <msItemStruct> requires careful consideration. In general, the focus of our markup was on the description of content and origin, while we paid less attention to such aspects of physical description as binding, which, at least in the 18th century, gives less information about the genesis of the manuscript than the binding of medieval manuscripts. Nevertheless, we encoded watermarks and other features that offered any recognizable temporal information relevant to questions of origin.

```

<category xml:id="nrss_aut_ecc">
  <catDesc>
    <term xml:lang="slv">ekleziasti&#x10d;ni</term>
    <term xml:lang="eng">ecclesiastic</term>
  </catDesc>

  <category xml:id="nrss_aut_ecc_sec">
    <catDesc>
      <term xml:lang="slv">škofijski kleriki</term>
      <term xml:lang="eng">secular clergy</term>
    </catDesc>
  </category>
<category xml:id="nrss_aut_ecc_mon">
  <catDesc>
    <term xml:lang="slv">redovništvo</term>
    <term xml:lang="eng">monastic orders</term>
  </catDesc>
<category xml:id="nrss_aut_ecc_mon_ofm">
  <catDesc>
    <term xml:lang="slv">fran&#x10d;iškani (OFM)</term>
    <term xml:lang="eng">Franciscan (OFM)</term>
  </catDesc>
</category>

```

Example 3. Another typology for sociocultural context of the manuscript authors was prepared in a similar way: **<taxonomy>** with nested **<category>** elements.

- 15 As to the content, we paid special attention to the classification of genre and the author's cultural background (see example 3). The retrieval software can use these pieces of explicitly encoded information and offer new insights, enabling the user to address new questions. Which textual genres prevailed in a particular area? Which social group has a prevailing share in some particular textual genre? Which geographical places are particularly bound to this or that kind of text?

3 Repository and Query System

- 16 For online searching and reading we have developed a portal based on the Fedora Commons server. Fedora Commons uses open standards to implement an XML-based digital document repository built from complex objects, their relationships, methods for transformation and presentation of data, and interfaces (Lagoze et al. 2006). The choice was based on the fact that the Fedora Commons server allows for construction of a repository conforming to the concept of an open archival information system with support for metadata harvesting.
- 17 Fedora Commons documents are expressed as XML documents and offer full support for namespaces. The documents are structured according to an object-oriented programming model. Each document has a number of data streams, analogous to internal object data,

and some repository metadata. The repository metadata are used for a number of important features of the Fedora Commons repository: a special data stream logs changes to a document and links to older versions of the object in the repository, a Dublin Core data stream stores document metadata for the repository search system and metadata presentation, and a Resource Description Framework (RDF) stream represents document relationships so that documents can not only be hierarchically organized but also have multiple arbitrary relationships according to the needs of a particular repository.

- 18 Access to all this rich data is implemented using a system of document disseminators, which are web-service access points and are very similar to class methods in an object-oriented system. Disseminators are implemented in special documents called content models (analogous to object classes), and each document can be associated with one or several content models. In this way, Fedora Commons can be regarded as an object-oriented system with no inheritance but where an object can still be part of more than one class. Disseminators are declared using WSDL (Web Service Definition Language) and use data streams and web service definitions to construct a specific presentation. In our case, disseminators are used to create HTML presentations of individual manuscript descriptions, biographical items, relationships, and collections using data streams from the underlying objects (such as TEI data, Dublin Core, XSLT stylesheets, RDF queries, or image data); even the facsimile browser is implemented as an XSLT transformation of a TEI `<facsimile>` element. This facility for creating web application interfaces in Fedora Commons has proven to be versatile since it permitted us not only to use different access methods on the XML documents in the repository to generate different presentations but also to combine the output of different disseminators to create advanced presentation features. Such chaining of XML data from one method to another, emulating what can be done in XML processing and presentation systems (i.e., Apache Cocoon), has been very helpful but occasionally resulted in performance penalties in the XSLT processing subsystem.
- 19 In the following subsections we shall present the components of the system and show how they were used to build the Register of Early Modern Slovenian Manuscripts repository.

3.1 Dublin Core Metadata

- 20 Each document in a Fedora Commons repository is identified not only by its permanent object identifier (PID) but also by its Dublin Core data stream, which is used for internal search. This data stream is directly available in the databases that support facilities such as object-relationship RDF query services, so extracting suitable Dublin Core metadata from the TEI document is a crucial step. For the Register, the conversion from TEI metadata, either in the header or in a particular `<msItemStruct>` or `<msPart>`, is problematic since TEI permits much richer encoding than Dublin Core, and there is no clear mapping to the Dublin Core tags.

3.2 Documents, Relationships, RDF

- 21 The repository uses RDF to express relationships between documents. RDF can be regarded as a very general method for storing data in the form of statements, where each statement is a triplet in the form of (subject, predicate, object), and so it is well suited to

representing inter-object relationships in this kind of a repository system. Fedora Commons also comes with a number of predefined predicates (or relationships), such as 'isMemberOfContainer', and supports several query languages for querying these relationships using the resource index web service interface provided by Fedora Commons. This interface is very useful since it is possible to define the template for the output, which can then be processed with normal XML processing tools, such as XSLT. An example is presented in example 4.

```

select $collTitle $collPid $member $creator $title
from   <#ri>
where  <info:fedora/nrss:bibl_064> <dc:title> $collTitle
and    <info:fedora/nrss:bibl_064> <dc:identifier> $collPid
and    $member <fedora-rels-ext:HasDescription> <info:fedora/
nrss:bibl_064>
and    $member <dc:creator> $creator
and    $member <dc:title> $title
or     $member <dc:identifier> 'nrss:bibl_064'
and    <info:fedora/nrss:bibl_064> <dc:identifier> $collPid

```

Example 4. An RDF query for all members of the repositories described in the bibliographic entry 064 (relation: 'HasDescription'), where Dublin Core elements <creator>, <title>, and <identifier> are extracted for each member. The resulting XML structure is converted into a listing as part of the transformation of a bibliographical entry. We store the document ID in the Dublin Core <identifier>, making it easy to produce HTML anchors in the list.

- 22 The document relationship system is a crucial part of the Fedora Commons repository, used mostly to organize the documents in different containers to represent collections. The RDF system also binds the documents to the project interface since the same Fedora Commons repository may contain digital objects related to other collections and projects. The fact that we use TEI for all projects in our Fedora Commons repository allows us to reuse code and keep a high level of standardization in the repository, but it is the RDF structure that holds the repository together.

3.3 Search Interfaces

- 23 Fedora Commons comes with a built-in search service, which can be used to search the internal metadata indices and Dublin Core fields of the documents in the repository. This system, while adequate for simple applications, was found insufficient for our purposes. Fedora Commons also supports several advanced full-text search plugins, so we have implemented our search infrastructure with the SolR plugin, giving us full power of the Lucene query engine and its indexing mechanism. Lucene can use XSLT to transform a document into a number of fields of data for indexing and then apply various plugins on the fielded data to implement features such as normalization, stemming, and removal of stopwords. Fedora Commons implements a tight integration with its search plugin interface, so it is possible to use its XML results in the disseminators to generate HTML result pages, allowing the users to treat the search system as just another built-in web service.

3.4 Putting It All Together

- 24 To create a web presentation and query system for the Register, we wrote a Perl script that transforms a TEI-encoded document into a collection of XML documents that can be imported as objects into the Fedora Commons server. We have created separate documents for each manuscript description or bibliographical item so that we can use the existing aggregation and collection facilities in the Fedora Commons repository and the mapping of Lucene-based search integration to repository objects. Then suitable disseminators for aggregation and presentation of textual items and their aggregates were developed using a number of custom-designed XSLT stylesheets. After the initial importing of TEI data into the system, all transformations are performed interactively using XSLT stylesheets as part of Fedora Commons disseminator specifications.
- 25 Each XML object, a manuscript description or bibliographical item, is displayed as one HTML page (an example is given in fig. 1). The site navigation bar contains a TEI option to download the complete TEI document, including the manuscript descriptions, the index of manuscripts, and the bibliographical database.

Figure 1. Rendering of the start of an `<msDesc>` element as an HTML page in the Fedora Commons repository using XSLT in the document disseminator. A search form field is shown at the top, and the "repository" field links to a search query for other manuscripts in the same archive. The navigation bar at the top has navigation links (next, previous, top of the archive, current collection, facsimile for the current manuscript), and export of its Dublin Core and TEI data. The content of the manuscript description gives Slovenian glosses for the TEI element names.

- 26 The HTML presentation to a large extent preserves the original structure of the TEI manuscript description or bibliographical item except that TEI element names are replaced with their Slovenian glosses and the document is transformed into HTML with suitable CSS styling (see fig. 1). The file giving the translation of TEI element glosses into Slovene is independently available for anyone wishing to use TEI-encoded data localized

to Slovenian. In addition, English glosses of TEI tags, attributes, and canonical values are provided, enabling an English rendering of the TEI structure in the HTML view as well. The system is expandable, providing for further localization as needed.

- 27 The output also takes advantage of the hyperlinked nature of the Register to directly display, for example, bibliographical items that refer to a manuscript in the scope of a <msDesc> element, or vice versa. The context links also enable users to click on the textual category of a manuscript (encoded as a cross-reference in the TEI) to browse other manuscripts in the same category. This feature takes advantage of the Fedora Commons facility for inter-object relationships to enable the portal to display further relationships using RDF queries in our disseminators. For example, where bibliographical items referring to a manuscript are enumerated in the manuscript description, a reverse lookup in the relationship database (expressed as a query to the RDF store in the system) is used to list all the mentioned manuscripts when the user views a bibliographical entry.

Figure 2. The facsimile viewer uses XSLT to create a web page from the <facsimile> element.

- 28 We have also developed a facsimile viewer, which enables viewing and browsing the manuscript images, as illustrated in figure 2. While the current version resembles an image browser, it is in fact an XSLT rendering of the <facsimile> structure into suitable HTML with CSS formatting. We are planning to expand its functionality by enabling ECMAScript-based features for more interactive paging, image pre-loading, zooming, and other features expected in an image browser while maintaining the relationship with the underlying TEI structure and still accommodating browsers that do not enable ECMAScript. As more transcriptions of the manuscripts in the Register become available, the facsimile viewer will be further enhanced to enable side-by-side viewing and other interactions between the facsimile data and available transcriptions.
- 29 The query system for the Register (built using the Fedora Commons SolR/Lucene plugin) has been developed to make it as easy as possible to run simple queries but also provides an advanced query form, where it is possible to search on multiple fields and modify the

search parameters. The user can also use the Lucene query language to input a custom query. Since all the fields from Dublin Core and from the TEI <msDescription> are presented to users, this allows for very powerful queries.

Figure 3. Sample search form in the Register.

- 30 We also use the same search infrastructure internally for querying the repository where the RDF store and its predefined inter-document relationships are not sufficient, which enables us to use many fields as HTML links to suitably formatted context lists, so that the users can view a list of all manuscripts in a particular archive or by a given hand with a single click (see figure 1 for an example).
- 31 Each object also provides links to Dublin Core elements describing it, which give standardized metadata access to the repository and enable metadata harvesting using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). In this way, the data are available to be included in aggregating resources and search engines like the CERL (Consortium of European Research Libraries) Portal, which uses OAI-PMH to aggregate the data of other libraries and resources (Mattheson 2005). So far, we have only integrated the repository with other repositories built on the same system (cf. Erjavec et al. 2011), but we have recently started participating in an initiative to create a national metadata harvester for digital repositories in humanities and expect to work towards inclusion in international indices. We hope this will improve the visibility of our work and will allow more researchers to benefit from the data available.

4 Temporal and Geographical Data

- 32 Our intention was that the Register would prove to be not only a decent manuscript catalogue but also a tool that will contribute to a more systematic account of these manuscripts in various dimensions. The data about the origin of the manuscripts can give us an insight into the temporal and territorial contexts ("chronotope") where the authors and genres appear: an insight into some peculiarities of time and space that shaped the manuscripts' content. In this way, we may notice more easily that a specific monastery, for instance, gave rise to an astonishing variety of texts in a given period, while, on the other hand, several vernacular writers in a relatively wide area copied and varied only a limited group of very specific texts. Spatiotemporal observations of this kind, supported by the query system, may open new insights. Besides their historical time and space, there is something that unites the text and the manuscript as a physical object—the "hand," the scribe, the person who is the source of the text—that may become an object of further research.

- 33 In this section we give an overview of the features of the Register that could support such analyses, by exemplifying their encoding, where we stress those aspects of the encoding which present problems for a fully automatic analysis.
- 34 The temporal dimension—the dating of the manuscripts—is presented in (Slovenian) prose and marked up as `msDesc/history/origin/origDate`, which carries attributes with ISO 8601 values for date representations. TEI offers several attributes to represent the time, as shown in the following three examples from the Register, where the prose has been, for these examples, translated into English:
1. `<origDate when="1659" cert="low">ca. 1659</origDate>`
 2. `<origDate notBefore="1775" notAfter="1825">late 18th or early 19th century</origDate>`
 3. `<origDate notBefore="1748-04-06" notAfter="1754-04-06">April 6th (probably 1753, not before 1748, not after 1754)</origDate>`
- 35 As can be seen, the dating of a manuscript can be either a particular date (year) or a time interval, where, additionally, confidence values can be assigned. These complications present problems for machine processing of the "chronos" aspects, as they require a complex temporal model to reason against.
- 36 The geographical dimension of the origin of the manuscripts—the "topos" aspect—is represented in `msDesc/history/origin/origPlace`, which contains the name of the place where the manuscript was written or used, as is shown in the following examples (also translated into English):
1. `<origPlace>Sv. Mihael by Novo Mesto</origPlace>`
 2. `<origPlace>Koroška</origPlace>`
 3. `<origPlace>Sveti Križ</origPlace>, <origPlace>Kranj</origPlace>, <origPlace>Trst</origPlace>`
 4. `<origPlace>Ljubljana or central Slovenia</origPlace>`
- 37 The first example specifies the village Sv. Mihael near the city of Novo Mesto. In 1979 the village was incorporated into Novo Mesto, and the area is today called Šmihel. The second example refers to Carinthia, a bilingual Slovenian-German speaking region, today in Austria. The third example gives a series of place names to indicate that the manuscript is a collection, originally written in several places. The last specifies the place of origin as the city of Ljubljana or the central region of Slovenia.
- 38 As with temporal information, formal modeling of the place names turns out to be a complex process. Ideally, the encoding and processing should support conjunction and disjunction of place names for a single object. The place names themselves should be given in their historical form, modern equivalent (if any), and translated into the name of the country in which the place is currently located (when appropriate), supplemented with a set of coordinates or links to geo-information portals. Adding this information and making use of it in the search system remains as further work.
- 39 Another feature that should be improved in the representation of our data is the relationship between literary genres and authors or their sociocultural contexts. Even now, the Register can be searched by combined criteria to give informative insights. For instance, if we look for manuscripts coming from the "peasant" sociocultural context, we are surprised to find among them some sermons of semi-apocryphal origin that have never been used in an ecclesiastical context but were used for private, domestic religious

reading in the mountain area of Carinthia. It is true that results of this kind could also be obtained by means of traditional research; still, the search tools in this Register make such interesting aspects much more evident and explicit.

- 40 We can envisage that the full potential of these "genre-plus-author" data would come into play even more in combination with more precise markup of the geospatial data, which would enable us to provide visualization tools. We plan to provide interactive web interfaces, where the visualization tools would use ECMA script to generate graphics and would use existing public interfaces, such as Google's mapping tools, for geodata presentation. TEI enables the expression of a rich set of distinctions, but the distinctions are usually ultimately expressed in prose, needing another layer of normalization to make them truly interoperable, as discussed above. Even if the information is formally encoded as ISO values of attributes, the attributes' complexity makes them difficult to model formally.

5. Conclusions

- 41 Although the Register of Early Modern Slovenian Manuscripts is not yet a comprehensive collection, it provides unrestricted access to many previously unknown or inaccessible manuscript texts from the Baroque and Enlightenment periods. There, the interested public—academic as well as general—can find texts which are scattered in many collections, public and private, in Slovenia and in the neighboring countries with Slovenian minorities (Austria and Italy). The retrieval system can perform searches against various criteria such as the place and date of origin of the manuscript or the holding institution.
- 42 The Register already enables export of its data. However, to make it fully interoperable, it would also need to be knit into the semantic web. While the availability of the TEI source, which is a standard format, the Dublin Core metadata, and the OAI-PMH service promote interoperability, a new feature of the Register, the implementation of which is currently in progress, will be the addition of a biography section, which will present the lives of the identified scribes and link them with the hand descriptions in the manuscript descriptions. This "atlas of hands" will enrich the Register by enabling a per-scribe analysis of the manuscripts—especially interesting in cases where the same person wrote a number of manuscripts, as it will make a comparison of the changes of writing style during their lifetime possible.
- 43 An important part of the Register which is currently missing is the addition of manuscript transcriptions. We are currently experimenting with this addition on some selected texts, but the majority of this effort is still in the planning stages. In the meantime, the TEI encoding of manuscript descriptions is being continuously improved: the more data are encoded, the more alive, visible, and "eloquent" the manuscripts can become for the audience.

BIBLIOGRAPHY

- Branschofsky, Margret, and Daniel Chudnov. 2002. "DSpace: Durable Digital Documents." In *Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries* (New York: ACM), p. 372. doi:10.1145/544220.544319.
- Burnard, Lou. 2005. "Encoding Standards for the Electronic Edition." In *Znanstvene izdaje in elektronski medij*, edited by Matija Ogrin, 25–42. Ljubljana: ZRC SAZU.
- Driscoll, Matthew James. 1997. *The Unwashed Children of Eve: The Production, Dissemination and Reception of Popular Literature in Post-Reformation Iceland*. London: Hisarlik Press.
- Erjavec, Tomaž, Jan Jona Javoršek, Matija Ogrin, and Petra Vide Ogrin. 2011. "Od biografskega leksikona do znanstveno-kritične izdaje: vprašanje trajnosti elektronskih besedil [From a Biographic Lexicon to a Text-Critical Edition: The Question of Durability of Digital Texts]." *Knjižnica: Revija za področje bibliotekarstva in informacijske znanosti* 55(1): 103–14.
- Kenney, Anne R., and Nancy Y. McGovern. 2003. "The Five Organizational Stages of Digital Preservation." In *Digital Libraries: A Vision for the 21st Century: A Festschrift in Honor of Wendy Lougee*, edited by Patricia Hodges, Mark Sandler, Maria Bonn, and John Price Wilkin, 122–53. Ann Arbor: The Scholarly Publishing Office, University of Michigan Library.
- Lagoze, Carl, Sandy Payette, Edwin Shin, and Chris Wilper. 2006. "Fedora: An Architecture for Complex Objects and Their Relationships." *International Journal on Digital Libraries* 6(2): 124–38. doi:10.1007/s00799-005-0130-3.
- Love, Harold. 1993. *Scribal Publication in Seventeenth-Century England*. Oxford: Oxford University Press.
- Matheson, Ann. 2005. "The Consortium of European Research Libraries (CERL): The European Printed Archive." *Alexandria: The Journal of National and International Library and Information Issues* 17, no. 1: 1–11.
- McKitterick, David. 2003. *Print, Manuscript, and the Search for Order, 1450–1830*. Cambridge: Cambridge University Press.
- Moureau, François, ed. 1993. *De bonne main. La communication manuscrite au XVIIIe siècle*. Paris: Universitas; Oxford: Voltaire Foundation.
- Ní Úrdail, Meidhbhín. 2000. *The Scribe in Eighteenth- and Nineteenth-Century Ireland: Motivations and Milieu*. Münster: Nodus Publikationen.
- Powell, Andy, and Pete Johnston. 2003. "Guidelines for Implementing Dublin Core in XML." Accessed February 23, 2012. <http://dublincore.org/documents/dc-xml-guidelines/>.
- Consultative Committee for Space Data Systems (CCSDS). 2012. *Reference Model for an Open Archival Information System (OAIS)*. Recommended Practice, Issue 2. Washington, DC: CCSDS Secretariat, NASA. June 2012, CCSDS Document 650.0-M-2. <http://public.ccsds.org/publications/archive/650x0m2.pdf>.
- TEI Consortium. 2012. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, edited by Lou Burnard and Syd Bauman. Version 2.0.2. Last updated February 2. N.p.: TEI Consortium. <http://www.tei-c.org/Guidelines/P5/>.

NOTES

1. <http://ezb.ijs.si/nrss/>
-

ABSTRACTS

This paper presents the Register of Early Modern Slovenian Manuscripts, which includes manuscripts from the 17th and 18th centuries that have been overlooked by scholars focused on printed books from the same era. The Register attempts to address this gap in Slovenian manuscript studies by describing these unknown and forgotten early modern manuscripts with facsimiles, an index of basic manuscript citations, and a bibliography containing publications where these manuscripts are mentioned. It is encoded in TEI P5 using the manuscript description module and available via a web interface. The paper discusses the specifics of early modern manuscripts, explains the structure and encoding of the Register (especially encoding of temporal and geographic data), and presents the portal built using the Fedora Commons repository software that allows the user to browse and search manuscripts and export data in TEI format, and that enables metadata harvesting.

INDEX

Keywords: Slovenian literature, codicology, manuscript description, early modern manuscripts, Fedora Commons

AUTHORS

MATIJA OGRIN

Matija Ogrin is a research fellow at the Research Centre of the Slovenian Academy of Sciences, Ljubljana, Slovenia. He combines traditional research in early modern Slovenian literature with methods for encoding and publication of scholarly digital editions of primary sources. His areas of interest include early modern sermon studies and manuscript description.

JAN JONA JAVORŠEK

Jan Jona Javoršek works at the Centre for Networking Infrastructure at the Jožef Stefan Institute, Ljubljana, Slovenia, where he is responsible for distributed networking computing infrastructure and development. He is also involved in research and development in the fields of language technologies, web interfaces, and public key infrastructure systems.

TOMAŽ ERJAVEC

Tomaž Erjavec works at the Department of Knowledge Technologies at the Jožef Stefan Institute in Ljubljana, Slovenia. His research interests are primarily in the field of language technologies, mostly in development of language resources, complex textual editions, and standardization of language encoding. Here he has been active in the work of TEI and ISO TC 37. Lately, his work has focused on developing corpora, lexica, and language models for processing historical Slovenian.