Manuscripts and old prints of the Karaim Digital Archive – from Ms Word files to TEI P5

Mariola Abkowicz[1], Tomasz Hoffmann[2], Michał Kozak[2]

[1]*Department of Asian Studies, Adam Mickiewicz University*
Aleja Niepodległości 24, 61-714 Poznań, Poland
E-mail: mariola.abkowicz@amu.edu.pl

[2]*Poznań Supercomputing and Networking Center*
Jana Pawła II 10, 61-139 Poznań, Poland
E-mail: {tomhof,mkozak}@man.poznan.pl

Abstract: Gathering metadata regarding historical resources, such as manuscripts and old prints, requires determining their structure. Archiving data in the form of Ms Word files containing descriptions of these objects seems to be a simple and effective solution. However, problems arise when we want to unify recurring information, analyze and publish metadata and accompanied scans to a wide range of researchers and the community. In that case, the TEI P5 format (http://www.tei-c.org/) can help us. We present the process of creating a digital archive of Karaim literature called *eJazyszłar*. We describe source data of this archive and justify the proposed solutions, especially TEI P5.

Key words: Karaim manuscripts, Karaim old prints, digital archives, TEI P5, eJazyszłar

I. INTRODUCTION

Monuments of Karaim literature from the 17–19th century are the only highly valuable source of historical knowledge about the Karaim language and its development. They are also an exceptional testimony to the Karaim spiritual culture. They consist of religious (translations of the Torah and the Psalms, prayers, religious songs), ritual (elegies), folklore (folk medicine, divination), and literary texts (including translations into Karaim of works by Jan Kochanowski – one of the first translations of the poet's works). They are preserved on the pages of manuscripts and old prints stored in a unique collection in Poland (one of the few in the world) of Karaim literature. These manuscripts, old prints and also publications from the early 20th century were inventoried within Grant DEC-2011/03/D/HS2/00618 of the National Science Centre

(Poland) led by Dr. Michał Németh. The result of this grant was a catalog of manuscripts and old prints written in Ms Word files. In effect the result had limited value for society and experts.

The catalog required a dedicated system with a database to systematize further inventory works and to provide browsing and searching for researchers interested in Karaim language. Work on such a system was partially financed by Grant 05570/15 of the Polish Ministry of Culture and National Heritage. Specialists of Poznań Supercomputing and Networking Center (PSNC), after familiarizing themselves with the gathered and developed metadata proposed to build on TEI P5 – the standard of representing texts, their semantics and metadata, which is becoming more and more popular among scientists in various fields of humanities [1, 2]. They implemented a parser of Ms Word files that transforms the metadata to the TEI P5 format and put them into the XML da-

tabase *eXistdb*[1]. Then they designed and created the website *eJazyszłar* (http://jazyszlar.karaimi.org) that operates on this database, shares its content in a user-friendly way, and allows logged-on editors to edit `<teiHeader>` via web forms.

In the next sections we describe the catalog, content of Ms Word files, the structure of our TEI P5 files (since TEI P5 is a quite generic format and some of the metadata can be placed in different ways) and the *eJazyszłar* system itself.

II. KARAIM MANUSCRIPTS AND OLD PRINTS

20th-century literature of Karaim mentions the existence of collections of Karaim manuscripts and old prints that are extremely valuable for the preservation of the cultural heritage of the Karaims and the study of their language. It urges researchers to take inventory of these documents, especially of the text remaining in private collections in Poland [3-5]. A significant number of these texts were not previously known to the community.

Completed works, in recent years, have focused on building a catalog of these manuscripts and old prints which have greatly expanded prior knowledge on Karaim literature [6-9]. Due to these works on establishing the place and time of their creation and the identity of their authors, translators and copyists, historians and linguists have gained new perspectives for their research [10].

Four hundred and sixty items of manuscripts and old prints stored in private collections in Poland were developed. For each of them a catalog sheet was created in an Ms Word file. The form of this sheet was developed according to tips for cataloging manuscripts in library collections [11] and the guidance for the development of manuscripts in Polish libraries [12]. We also took into account the specific nature of the handwritten documents and the handwritten cards bound into the printed documents. In thirty cells of the sheet, the following metadata was registered: the creators with their responsibilities (authors, translators, copyists, etc.), languages with dialects, titles (determined precise or descriptive), the type of document, place and date of issue. In the case of the old prints the publishers and printing houses were recorded. In the case of the manuscripts the type of alphabet and the style of letters were recorded. Important information was whether the text was covered, what type of cover, paper and ink were used, Other significant metadata included the number of pages, the number of handwritten cards and printed cards, block and card format in millimeters and the current condition of the document and its cover.

A particularly important part of the sheet was the cell with the information about the content. This cell was very elaborate, especially for documents that consist of many handwritten cards created in different places, at different times and written by different authors, translators and copyists. Besides this information, for each item we also recorded the literary genre, the language with dialects, glosses in other languages (Hebrew, Russian, Polish and Karaim dialects), the incipit and excipit and remarks.

Tab. 1 presents the part of the sheet for manuscript http://jazyszlar.karaimi.org/manuscript/Man.432.

III. TEI P5

As we mentioned in the introduction, we decided to use the TEI P5 standard as a metadata container. There are three main factors that support this option.

Firstly, the TEI P5 header is so rich in detail that almost each of the metadata from Ms Word files finds its own place. Even complex items of information about the content have their counterparts in this header. Other descriptive metadata standards commonly used in digital libraries, like DCMI Terms[2], MARC 21[3] and MODS[4], do not have such fields. There is also a lack of such fields in administrative metadata: METS[5] and PREMIS[6]. On the other hand, even if a field from an Ms Word file has no direct counterpart in TEI P5, the standard allows for the usage of elements and attributes in a very flexible way. We refer readers not familiar with TEI P5 to its guidelines [13].

Secondly, TEI P5 files as XML files are a perfect data model. We do not need to put the metadata into a relational database to query and update them. We can use one of the available XML databases and take advantage of its XQuery API [14].

Thirdly, the TEI P5 standard allows us to attach scans of a manuscript and many text layers. In this way it allows editors to express the syntax and semantics of a text in a very flexible way, ranging from general statements on the textual structure to highly specific analyses of all kinds of textual phenomena. We plan to take advantage of these features in the future. Currently *eJazyszłar* uses the `<text>` element only in a basic manner; to make references to scans of pages (where they are accessible) and to provide the incipit and excipit of a document. For the `<teiHeader>` element, almost all its sub-elements are handled:
- `<titleStmt>`,
- `<publicationStmt>`,
- `<biblStruct>`,

---

[1] http://exist-db.org/
[2] http://dublincore.org/documents/dcmi-terms/
[3] http://www.loc.gov/marc/
[4] http://www.loc.gov/standards/mods/
[5] http://www.loc.gov/standards/mets/
[6] http://www.loc.gov/standards/premis/

Tab. 1. A fragment of the structure of Ms Word file

Description's author	Michał Németh, Anna Sulimowicz
Inventory number	JSul.III.30
Title / Name	סדור התפלות כמנהג הקראים [*Siddur ha-tefilot ke-minhag ha-qara'im*]
Print / Manuscript	Druk ‖ Rękopis (wstawki rękopiśmienne)
Presence of handwritten cards	Tak
Publication date / Creation date	1836 ‖ poł. XIX w
Publication place / Creation place	Eupatoria (Gözleve) ‖ Halicz
Author	— ‖ Niepodany
Translator	— ‖ Zob. *Informacja o treści*
Publisher	Mordechaj Tiriszkan ‖ —
Copyist	— ‖ Jeszua Josef Mordkowicz (1802–1884)
Printing house	Drukarnia Mordechaja Tiriszkana ‖ —
Information about content	Druk: modlitwy na dni świąteczne. Wstawki rękopiśmienne: **1)** 1 rᵒ – 2 vᵒ: przekład hymnu na język karaimski (karty wszyte między 6 a 7 k. druku) incipit: אומהסי ישראלנין ציקקנלרינדא מיצרידין *Ummasy Israelnin cykkanlarynda Micriden* dialekt: południowozachodni autor: niepodany tłumacz: Josef syn Abrahama (Leonowicz) **2)** 5 rᵒ – 5 vᵒ: przekład na język karaimski hymnu religijnego (pijut) na święto Szawuot (karta wszyta między 98 a 99 k. druku) incipit: אדני טנרישי ישראלנין *Adonaj Tenrisi Israelnin* dialekt: południowozachodni autor: niepodany tłumacz: kopista, Jeszua Josef syn Moszego (Mordkowicz) **3)** 6 rᵒ – 6 vᵒ: przekład na język karaimski hymnu religijnego (pijut) na święto Szawuot (karta wszyta między 101 a 102 k. druku) incipit: אדני בי ישראלנין *Adonaj biji Israelnin* dialekt: południowozachodni autor: niepodany tłumacz: szwagier kopisty, Josef syn hazzana Abrahama (Leonowicz)
Language	Hebrajski ‖ Hebrajski, karaimski
Karaim dialect	Południowozachodni
Glosses	—
Letters	Hebrajskie
...	...

- `<msDesc>`,
- `<physDesc>`,
- `<profileDesc>`.

In Appendix we list TEI P5 obtained from Table 1.

IV. eJazyszłar

Let us start this section with a diagram presenting the process of populating a database with initial data and the architecture of *eJazyszłar*. We suggest reading this diagram from the lower left corner.
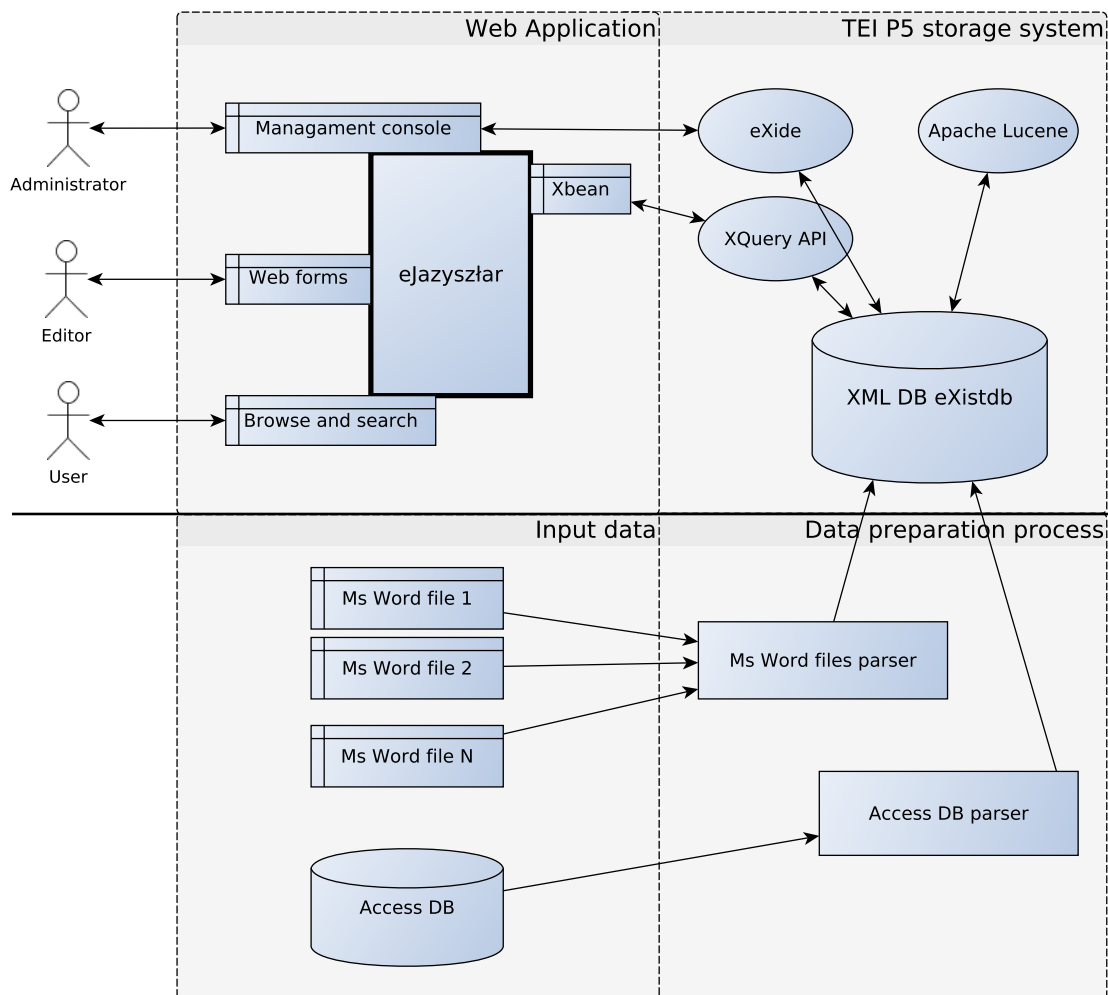
Fig. 1. Populating *eJazyszłar* with initial data and its architecture

We had two sources of input data: Ms Word files and data from an Ms Access database. We will focus on Ms Word files since it is the main topic of this paper.

As is seen in Tab. 1, we had to parse not only a table, but also cells. In many cells the sign ‖ separates values for an old print from values for a manuscript. Moreover, these values can be complex, i.e. many names are listed as authors, translators, publishers, copyists (usually after the comma), titles can have their transliteration written in italics, languages can be optionally specified by dialects, etc. However the most difficult part for our parser was the cell with information about content (13th row in Tab. 1). There are many non-mandatory parts of items: locus (a page or a page range), title, incipit with its transliteration (in italics), dialect, authors, translators and remarks. Moreover, we had to recognize which tokens are names of people, in order to make references in TEI P5 (in the `sameAs` attribute, e.g. in line 85 of Appendix) and which are only descriptions of people (for instance we have

"the copyist's brother-in-law" in line 111).

We implemented our parser in Java 8 with the usage of JAXB[7]. It means we generated the structure of Java classes for TEI P5 from its schema, and in runtime we filled that structure with parsed data from Ms Word files. JAXB marshalling yielded valid TEI P5 files which we next uploaded to the open source native XML database *eXistdb*, that is used as a storage system of *eJazyszłar*.

Regarding the parser of data from Access DB, we would only like to mention that this database contains the data of Karaim literary bibliography. In comparison to the way the data of the Ms Word files was structured, implementing the parser was easier. The only challenge was merging people identifiers from these two data sources (see, for instance, Aleksander Mardkowicz's bio http://jazyszlar.karaimi.org/person/Per.28).

Let us go now to the upper left corner of the diagram. *eJazyszłar* is a web application built according to responsive web design principles, i.e. it automatically expands, rearranges,

---

contracts and removes content on the basis of the screen size. It functions flawlessly on different browsers and devices like PCs, laptops, tablets, and smartphones.

According to the rights of the logged-in user, the system provides different functionalities. Users that are not logged-in can browse and search collections. It is worth mentioning that *eXistdb* supports the concept of XML collections. These collections allow us, among other things, to query only one of them, and define which elements in TEI P5 should be indexed. We are taking advantage of these features since TEI P5 of manuscripts and old prints, books, chapters, journals, articles, etc., differ from one another in some aspects. Collections are reflected in the menu of *eJazyszłar* (see online) and are used in search filters as well. *eJazyszłar* provides full text searching based on *Apache Lucene*[8] – an ultra-fast search engine, tightly integrated with *eXistdb* [15].

Logged-in users are editors of TEI P5 files. However they do not need to know their structure since *eJazyszłar* provides user-friendly web forms through which they can create and modify data. The most complex web forms are those for the manuscript and old print collection (a part of this form is shown in Appendix). They consist of four steps that cover all subelements of `<teiHeader>` listed in Section III. In the fourth step editors can attach scans of manuscripts and old prints as well. They are attached in `<body>` as pages with the `facs` attribute. *eJazyszłar* presents these files in the form of a web gallery with a zoom in function.

The most privileged group of editors are the administrators of *eJazyszłar*. They have an additional web form to manage the accounts of the editors and have access to *eXide* – a web module of *eXistdb* that allows logged in users to edit XML files stored in *eXistdb*. This web console also allows them to write and execute queries in the XQuery language.

Besides *eXide* as an interface to query and modify XML data, *eXistdb* provides a REST-style interface. Queries written in the XQuery language can be sent to the database by the POST method of the HTTP protocol. This XQuery API was attractive in the context of designing the *eJazyszłar* architecture. We built our system as a multi-tier web application with a data access layer using this interface. We coded in Java 8 and Play Framework[9], which comes with Ebean[10] as a library for accessing data in relational databases. However we did not have here a relational database but an XML database. Following the Ebean implementation we implemented a submodule of *eJazyszłar* (called Xbean) that allowed us to base our data access layer on it. For instance, with the usage of Xbean, a Java method of returning a list of TEI P5 objects with bios of people whose names contain a given string looks as follows:

```java
public TEIList getPeopleByName(final String name) {
  return Xbean.newXQuery(TEIList.class)
    .setHowMany(10)
    .where()
     .disjunction()
       .contains("/tei:text/tei:body/tei:listPerson/tei:person/tei:persName/tei:forename", name)
       .contains("/tei:text/tei:body/tei:listPerson/tei:person/tei:persName/tei:surname", name)
     .endJunction()
    .order().asc("/tei:text/tei:body/tei:listPerson/tei:person/tei:persName/tei:surname")
    .findList("people");
}
```

The term "people" in line 10 is the name of a TEI P5 collection with bios. Execution of this method with the parameter "Mor" sends the query below to *eXistdb* and returns an instance of the TEIList class. This class is a wrapper around a list of TEI classes that were generated by JAXB tools (as we mentioned above). As a result of the execution we obtain a list of parsed TEI P5 objects that we can arbitrarily process in Java, render on a web page or a web form.

```xml
<query xmlns="http://exist.sourceforge.net/NS/exist" start="1" max="10" cache="no">
  <text>
    xquery version "3.0" encoding "utf-8";
    declare default collation "?lang=pl-PL";
    declare namespace tei="http://www.tei-c.org/ns/1.0";
    for $x in /tei:tei
    where
      (
        $x/tei:text/tei:body/tei:listPerson/tei:person/tei:persName/tei:forename[contains(lower-case(.),
          lower-case('Mor'))]
        or
        $x/tei:text/tei:body/tei:listPerson/tei:person/tei:persName/tei:surname[contains(lower-case(.),
          lower-case('Mor'))]
      )
    order by $x/tei:text/tei:body/tei:listPerson/tei:person/tei:persName/tei:surname ascending
    return $x
  </text>
</query>
```

---

[8] http://lucene.apache.org/

[9] http://www.playframework.com/

[10] http://ebean-orm.github.io/

This example of a query is rather simple, but Xbean allows us to create quite complex queries by joining "where" and "order" clauses in chains with the usage of Boolean operators.

V. CONCLUDING REMARKS AND FUTURE WORKS

eJazyszłar is a digital archive of Karaim literature in which metadata is written and stored in TEI P5 format. This system can be adjusted for other uses and for other bibliographies. PSNC can expand the set of TEI P5 elements handled by the system and develop new system skins as well. Therefore, on the one hand this system can be an alternative to digital library systems (but based on TEI P5), and on the other hand, it can be an option for digital humanists dealing with digital scholarly editions in TEI P5 [16].

Currently our system is not perfected for any of these applications. For instance, there is a lack of the OAI–PMH protocol[11] a lack of support for delivering images according to the IIIF specification,[12] and a lack of support for marking texts in many dimensions. As we mentioned in Section III., future works will focus on the last topic. Nevertheless, the current version of the system is quite good for further development in both directions. We encourage Readers to contact the authors for cooperation.

Acknowledgements

TEI P5 of `http://jazyszlar.karaimi.org/manuscript/Man.432`

```xml
<?xml version="1.0" encoding="utf8"?>
<tei:tei xmlns:tei="http://www.tei-c.org/ns/1.0">
  <tei:teiHeader>
    <tei:fileDesc>
      <tei:titleStmt>
        <tei:title xml:lang="heb-Hebr">סדור התפלות כמנהג הקראים</tei:title>
        <tei:principal>
          <tei:persName sameAs="Per.115">Michał Németh</tei:persName>
        </tei:principal>
        <tei:principal>
          <tei:persName sameAs="Per.47">Anna Sulimowicz</tei:persName>
        </tei:principal>
      </tei:titleStmt>
      <tei:publicationStmt>
        <tei:publisher>
          <tei:orgName>Związek Karaimów Polskich</tei:orgName>
        </tei:publisher>
      </tei:publicationStmt>
      <tei:sourceDesc>
        <tei:biblStruct type="print">
          <tei:monogr>
            <tei:title xml:lang="heb-Hebr" type="main">סדור התפלות כמנהג הקראים</tei:title>
            <tei:title type="transliteration">Sidur ha-tefilot ke-minhag ha-qara'im</tei:title>
            <tei:author>
              <tei:persName>--</tei:persName>
            </tei:author>
            <tei:editor role="publisher">
              <tei:persName sameAs="Per.684">Mordechaj Tiriszkan</tei:persName>
            </tei:editor>
            <tei:editor role="translator">
              <tei:persName>--</tei:persName>
            </tei:editor>
            <tei:editor role="copyist">
              <tei:persName>--</tei:persName>
            </tei:editor>
            <tei:textLang mainLang="heb-Hebr"/>
            <tei:imprint>
              <tei:pubPlace>Eupatoria (Gözleve)</tei:pubPlace>
              <tei:date when="1836"/>
              <tei:respStmt>
```

```xml
41        <tei:resp>printing</tei:resp>
42        <tei:name>Drukarnia Mordechaja Tiriszkana</tei:name>
43      </tei:respStmt>
44     </tei:imprint>
45    </tei:monogr>
46   </tei:biblStruct>
47   <tei:biblStruct type="manuscript">
48    <tei:monogr>
49     <tei:author>
50      <tei:persName>Niepodany</tei:persName>
51     </tei:author>
52     <tei:editor role="publisher">
53      <tei:persName>--</tei:persName>
54     </tei:editor>
55     <tei:editor role="translator">
56      <tei:persName>Zob. Informacja o treści</tei:persName>
57     </tei:editor>
58     <tei:editor role="copyist">
59      <tei:persName sameAs="Per.384">Jeszua Josef Mordkowicz (1802-1884)</tei:persName>
60     </tei:editor>
61     <tei:textLang mainLang="kdr-gal-Hebr"/>
62     <tei:textLang mainLang="heb-Hebr"/>
63     <tei:imprint>
64      <tei:pubPlace>Halicz</tei:pubPlace>
65      <tei:date when="poł. XIX w."/>
66     </tei:imprint>
67    </tei:monogr>
68   </tei:biblStruct>
69   <tei:msDesc>
70    <tei:msIdentifier>
71     <tei:repository>Baza danych o rękopisach i starodrukach karaimskich w polskich zbiorach
       ↪   prywatnych</tei:repository>
72     <tei:idno>Man.432</tei:idno>
73     <tei:msName>JSul.III.30</tei:msName>
74    </tei:msIdentifier>
75    <tei:msContents>
76     <tei:summary>Druk: modlitwy na dni świąteczne. Wstawki rękopiśmienne:</tei:summary>
77     <tei:msItem n="1">
78      <tei:locus from="1 r" to="2 v"/>
79      <tei:incipit>אומהסי ישראלנין ציקקנלרינדא מיצרידין</tei:incipit>
80      <tei:incipit type="transliteration">Ummasy Israelnin cykkanlarynda Micriden</tei:incipit>
81      <tei:author>
82       <tei:persName>niepodany</tei:persName>
83      </tei:author>
84      <tei:editor role="translator">
85       <tei:persName sameAs="Per.503">Josef syn Abrahama (Leonowicz)</tei:persName>
86      </tei:editor>
87      <tei:textLang mainLang="kdr-gal-Hebr"/>
88      <tei:title>przekład hymnu na język karaimski (karty wszyte między 6 a 7 k. druku)</tei:title>
89     </tei:msItem>
90     <tei:msItem n="2">
91      <tei:locus from="5 r" to="5 v"/>
92      <tei:incipit>אדני טנרישי ישראלנין</tei:incipit>
93      <tei:incipit type="transliteration">Adonaj Tenrisi Israelnin</tei:incipit>
94      <tei:author>
95       <tei:persName>niepodany</tei:persName>
96      </tei:author>
97      <tei:editor role="translator">
98       <tei:persName>kopista</tei:persName>
99      </tei:editor>
100     <tei:textLang mainLang="kdr-gal-Hebr"/>
101     <tei:title>przekład na język karaimski hymnu religijnego (pijut) na święto Szawuot (karta wszyta
       ↪   między 98 a 99 k. druku)</tei:title>
102    </tei:msItem>
103    <tei:msItem n="3">
104     <tei:locus from="6 r" to="6 v"/>
105     <tei:incipit>אדני בי ישראלנין</tei:incipit>
106     <tei:incipit type="transliteration">Adonaj biji Israelnin</tei:incipit>
107     <tei:author>
108      <tei:persName>niepodany</tei:persName>
109     </tei:author>
110     <tei:editor role="translator">
111      <tei:persName>szwagier kopisty</tei:persName>
112     </tei:editor>
113     <tei:textLang mainLang="kdr-gal-Hebr"/>
114     <tei:title>przekład na język karaimski hymnu religijnego (pijut) na święto Szawuot (karta wszyta
       ↪   między 101 a 102 k. druku)</tei:title>
115    </tei:msItem>
116   </tei:msContents>
117   <tei:physDesc>
118    <tei:objectDesc>
```

```xml
119                    <tei:supportDesc material="paper">
120                      <tei:support>Papier czerpany, żeberkowy, w kolorze kremowym (wstawka nr 1) oraz gładki,
            ↪   nieczerpany (wstawka nr 2, 3), ślady liniowania za pomocą tzw.    [suralal]</tei:support>
121                      <tei:extent>
122                        <tei:measure unit="page">176 str. druku + 2 nlb. str. druku + 6 kk. rkp.</tei:measure>
123                        <tei:note>Numeracja: karty druku 1-144, 1-32, karty rkp.: brak numeracji</tei:note>
124                        <tei:dimensions type="block" unit="mm">
125                          <tei:width>265</tei:width>
126                          <tei:height>190</tei:height>
127                        </tei:dimensions>
128                        <tei:dimensions type="leaf" unit="mm">
129                          <tei:width>220</tei:width>
130                          <tei:height>175</tei:height>
131                        </tei:dimensions>
132                        <tei:dimensions type="leaf" unit="mm">
133                          <tei:width>250</tei:width>
134                          <tei:height>180</tei:height>
135                        </tei:dimensions>
136                        <tei:dimensions type="leaf" unit="mm">
137                          <tei:width>250</tei:width>
138                          <tei:height>180</tei:height>
139                        </tei:dimensions>
140                      </tei:extent>
141                      <tei:condition>Dobry</tei:condition>
142                    </tei:supportDesc>
143                    <tei:layoutDesc>
144                      <tei:layout columns="1" ruledLines="26 20"/>
145                      <tei:layout columns="1" ruledLines="35 22"/>
146                      <tei:layout columns="1" ruledLines="35 24"/>
147                    </tei:layoutDesc>
148                  </tei:objectDesc>
149                  <tei:typeDesc>
150                    <tei:typeNote medium="ink">Galusowy, ciemnobrązowy, czarny</tei:typeNote>
151                  </tei:typeDesc>
152                  <tei:scriptDesc>
153                    <tei:scriptNote script="NON-VOCALIZED"/>
154                    <tei:scriptNote script="Square Semi-cursive Cursive"/>
155                  </tei:scriptDesc>
156                  <tei:additions>
157                    <tei:p>Tom drugi modlitewnika wydanego w Eupatorii staraniem kupca Mordechaja Tiriszkana. Karty rkp.
            ↪   3 r o – 4 v o puste. Kopistę wstawek rękopiśmiennych według informacji zawartej w nagłówkach
            ↪   był Jeszua Josef Mordkowicz (1802-1884), tłumacza wstawki nr 2. Powstały one w drugiej połowie
            ↪   XIX wieku (nie później niż 1884), a według nagłówka utworu we wstawce nr 2 kopista dokonał
            ↪   tłumaczenia w młodości.</tei:p>
158                    <tei:p>Tom drukowany zawiera wkładki rękopiśmiennicze.</tei:p>
159                  </tei:additions>
160                  <tei:bindingDesc>
161                    <tei:condition>Drewno obleczone skórą, grzbiet zachowany, w kolorze brązowym, ślady żerowania
            ↪   owadów, blok zwarty, ozdobne wytłoczenia i napis na okładce o ofiarowaniu książki gminie
            ↪   Halickiej</tei:condition>
162                  </tei:bindingDesc>
163                </tei:physDesc>
164              </tei:msDesc>
165            </tei:sourceDesc>
166          </tei:fileDesc>
167          <tei:profileDesc>
168            <tei:textClass>
169              <tei:keywords>
170              </tei:keywords>
171            </tei:textClass>
172          </tei:profileDesc>
173        </tei:teiHeader>
174        <tei:text>
175          <tei:body>
176            <tei:p n="begin">--</tei:p>
177            <tei:p n="end">--</tei:p>
178          </tei:body>
179        </tei:text>
180      </tei:tei>
```

A part of the web form for the manuscript and old print collection

References

[1] A.H. Renear, *Text Encoding*, [In:] Susan Schreibman, Ray Siemens, and John Unsworth, editors, *A Companion to Digital Humanities* Oxford: Blackwell, 2004.

[2] J. Cummings, *The Text Encoding Initiative and the Study of Literature*, [In:] S. Schreibman, R. Siemens, eds., *A Companion to Digital Literary Studies* Oxford: Blackwell, 2008.

[3] T. Kowalski, *Najpilniejsze zadania karaimoznawstwa*, Myśl Karaimska, **11**, 12–23, (1935/1936).

[4] A. Dubiński, *Karaimische Handschriften in polnischen Sammlungen*, Central Asiatic Journal **23**(3–4), 147–150, (1979).

[5] A. Dubiński, Karaimskie rukopisi iz kollekcii Ju. Sulimoviča v Varšave, [In:] G.A. Abdurachmanov and A. P. Chodžiev, editors, *Tjurkskoe jazykoznanie. Materialy III vsesojuznoj tjurkologičeskoj konferencii*, pages 20–24, Taškent 10–12.IX.1980, 1985.

[6] M.Németh, *Karaim literature as a source of information on the spoken language. A case study of the early 20th-century Lutsk Karaim dialect*, Karaite Archives, **1**, 113–132, (2013).

[7] A. Sulimowicz, *Crimean Karaim manuscripts in the Józef Sulimowicz collection*, Rocznik Orientalistyczny(2), 169–179, (2015).

[8] A. Sulimowicz, *Lutsk Karaite manuscripts and documents in the Józef Sulimowicz collection*, Acta Orientalia Academiae Scientiarum Hungaricae **68**(2), 175–182, (2015).

[9] A. Sulimowicz, *XIX–wieczny rękopis z polskich archiwów prywatnych zawierający dwa dokumenty dotyczące Karaimów trockich*, Almanach Karaimski, **5**, 189–205, (2016).

[10] M. Németh, *Rękopisy i druki karaimskie w polskich zbiorach prywatnych. Nowe perspektywy badań karaimoznawczych*, Almanach Karaimski, **5**, 61–101, (2016).

[11] Polska Akademia Umiejętności, *Wskazówki do katalogowania rękopisów w zbiorach bibliotecznych*, Księgarnia Gebethnera i Wolffa, Kraków, 1935.

[12] B. Horodyski, H. Więckowska, *Wytyczne opracowania rękopisów w bibliotekach polskich*, Zakład imienia Ossolińskich, Wrocław, 1955.

[13] Text Encoding Initiative Consortium, *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, 2016.

[14] J. Robie, D. Chamberlin, M. Dyck, J. Snelson, eds., *XQuery 3.0: An XML Query Language*, W3C Recommendation, 2014.

[15] A. Retter, E. Siegel, *eXist: A NoSQL Document Database and Application Platform*, O'Reilly Media, 2014.

[16] S. Hockey, *Electronic Texts in the Humanities: Principles and Practice*, Oxford University Press, 2000.

Mariola Abkowicz, librarian, academic teacher, social activist. Graduated from the University of Wroclaw – Institute of Library Science and Information. Head of the Library at the Faculty of Dentistry of the Wroclaw Medical University. Associated with the Section of Hebrew, Aramaic and Karaim Studies, Department of Asian Studies at the A. Mickiewicz University in Poznań as a teacher of Karaim science, Karaim literature and culture. She is actively promoting the culture and heritage of Karaim and the knowledge about the multicultural history of Wroclaw. Since 1998 the President of the Board of the Polish Karaim Association. Publisher and editor-in-chief of the Karaim historical, socio-cultural quarterly "Awazymyz", "Almanach Karaimski", "e-Jazyszlar. The Karaim Digital Archive", "the Polish Karaim Publishing House Bitik" and the Internet portal www.karaimi.org. Editor of books and author of articles related to Karaims, and lecturer at popular science and scientific meetings in Poland and abroad. As a representative of the Karaim community, she participates in the work of the Joint Government and National Minorities Commission and the Kaleidoscope of Cultures.

Tomasz Hoffmann – in 2010 he received his MSc degree in Computer Science and BSc degree in Mathematics at the Poznan University of Technology, and currently he is finishing his PhD thesis in Computer Science. Since 2014 he has been working at Poznań Supercomputing and Network Center as a member of the Knowledge Platforms Team. His main areas of interest are: numerical methods, machine learning and information retrieval, in particular for digital humanities.

Michał Kozak, a university graduate in Computer Science and Doctor of Philosophy in Computer Science. He defended his PhD dissertation in computational logic in 2011 at the Faculty of Mathematics and Computer Science of the A. Mickiewicz University in Poznań. In the meantime, he has programmed in Java and C++ – among other things he co-created the XRayan application; he has also gained experience in the Banking Projects Polsoft company. Since 2010 he has been working for Poznań Supercomputing and Networking Center, initially as a designer and Java and Python programmer in the Digital Libraries Team, and currently in as leader of the Knowledge Platforms Team. Together with the team he creates and develops specialized systems and tools for e-humanities. He is author and co-author of over a dozen of research and conference papers on logic and IT systems.