

Information Technology and Digital Humanities Workflow in the Rubensohn Project

Research website and rich TEI XML header encoding

Daniel A. Werning (Berlin)

§1 Task

As part of the project “Die ägyptische und orientalische ‘Rubensohn-Bibliothek’. 4000 Jahre Kulturgeschichte einer altägyptischen Insel”, (hereafter “Rubensohn Project”, led by Prof. Dr. Verena M. Lepper (SMB, ÄMP), data from approximately one thousand inscribed objects are to be collected, digitally archived, and published on a research-oriented website. The artefacts are of different kinds, for example, papyri, stone ostraca, ceramic ostraca, and wooden tablets. They are inscribed with different scripts and languages: Aramaic, Greek, Coptic, Demotic, Hieratic Egyptian, and Hieroglyphic Egyptian. But they all come from Elephantine island.

Each of the above mentioned tasks demands different Information Technology (IT) techniques. Firstly, the digital archiving of data calls for sustaining text encoding standards such as Unicode and eXtensible Markup Language (XML) compliant encoding,¹ as well as sustaining high-quality image formats such as TIFF. We chose to also offer the data encoded in accordance to the guidelines of the “Text Encoding Initiative” (TEI).² On the other hand, the data collection has to meet the needs and customs of the contributing philologists, which are generally not very familiar with these IT techniques and which are locally distributed across Europe without necessarily permanent internet access. Moreover, it is preferable that the encoded data content is compliant to standardized classification thesauri as far as possible, to the end that cross-references or even data exchange between similar projects is easily possible. Finally, for the publication on an interactive website that enables the user to research the data and view the artefact images, the data have to conform to current online data formats such as (X)HTML, MySQL databases, and compact JPEG format, respectively. The website users shall also have the possibility to comment on information that they find on the webpages, for example to hint at mistakes or provide additional information.

§2 Data types and data formats

We collect three types of data for each artefact:

- i) *Images* of all, including empty surfaces (if possible).

1 Unicode Consortium, The Unicode Standard, v6.2.0, 2012, <<http://www.unicode.org/versions/Unicode6.2.0/>> (30.08.2014); World Wide Web Consortium, Extensible Markup Language (XML), <<http://www.w3.org/XML/>> (both 11.03.2013). A good introduction is Helmut Vonhoege, *Einstieg in XML: Grundlagen, Praxis, Referenz*, 6th ed., Bonn 2011.

2 See Text Encoding Initiative. P5 Guidelines, v2.2.0, 25.08.2012 <<http://www.tei-c.org/Vault/P5/2.2.0/doc/tei-p5-doc/en/html/index.html>> (30.08.2014).

- ii) Detailed *metadata*, i.e. information concerning the artefact like its inventory number(s), its size, the script and language of its inscription, and many more — altogether more than 100 pieces of information per artefact (see appendix below).
- iii) An *encoding* or *transliteration*, respectively, of the text that is inscribed on it, plus a *translation*.

The images were taken using a large size scanner produced by Cruse Spezialmaschinen GmbH. The images come with a fixed resolution of 600 dpi and specified color profile and they were saved in TIFF format. The pictures of certain artefacts, for example those that are fixated under glass, had to be taken by hand with a regular digital camera. For the online publication, these image files are enriched with copyright information and stored in JPG format. Besides the full resolution image, a medium sized preview version and a thumbnail version were computed.

The set of metadata to be collected was inspired, on the one hand, by a comparison of already existing online databases that provide metadata for ancient manuscripts³ and, on the other hand, by a comparison of different metadata tagging standards.⁴ Among them, the set of metadata in the “Berliner Papyrusdatenbank” project, the “Papyrus Projekt: Halle, Jena, Leipzig”, and the structure of metadata in the TEI header probably had the most influence. For example, the TEI guidelines for the <teiHeader> induce a strict separation between “origin”, “provenance”, and “acquisition” of the objects,⁵ and the IT architecture in the Papyrus Projekt suggests a strict separation between text-related data and support-related data.⁶ Of course, data fields specific to the Rubensohn Project was also included (e.g. information on the boxes that the artefacts were shipped in to Europe).

-
- 3 BerlPap. Berliner Papyrusdatenbank, <<http://ww2.smb.museum/berlpap/>>, 21.12.2012 (with many thanks for sharing their internal data; 30.08.2014); Papyrus und Ostraka Projekt Halle · Jena · Leipzig, documentation: Reinhold Scholl, Jens Kupferschmidt, Andy Wermke, et. al., Papyrus Projekt. Umsetzung digitaler Papyrus-Sammlungen auf Basis der Content Management Anwendung des MyCoRe Arbeitskreises, v4.5.3, Leipzig 21.11.2012, <http://papyri.uni-leipzig.de/Papyri_Konzept.pdf?XSL.lastPage.SESSION=/Papyri_Konzept.pdf> (30.08.2014); Stefan Freitag, Marius Gerhardt, Jens Kupferschmidt, Reinhold Scholl, Papyrus Portal. Umsetzung einer Portallösung für alle Papyrussammlungen in Deutschland, v1.4, Leipzig 15.03.2012, <http://www.papyrusportal.de/PapyrusPortal_Dokumentation.pdf> (15.12.2012); APIS Partner Data Element Comparison, <<http://www.columbia.edu/cu/libraries/inside/projects/apis/documentation/meta.comparison.html>> (22.11.2012); Trismegistos. Specific Texts, <<http://www.trismegistos.org/tm/search.php>> (19.11.2012); Heidelberger Gesamtverzeichnis der griechischen Papyrusurkunden Ägyptens. Hauptregister, <http://aquila.papy.uni-heidelberg.de/Hauptregister/FMPPro?-db=hauptregister_-&-format=DSearch.htm&-lay=Suche&-max=1&-token=25&-view> (15.12.2012); Stephan J. Seidlmayer, Ingelore Hafemann, Handbuch zur Benutzung des Thesaurus Linguae Aegyptiae (TLA), Berlin 2011, <http://aaew.bbaw.de/hgl/0/Manual_2012_02_02.pdf> (with many thanks for also sharing their internal data; 15.12.2012); Deir el Medine online, <<http://dem-online.gwi.uni-muenchen.de>> (21.11.2012).
 - 4 TEI. P5: Guidelines for Electronic Text Encoding and Interchange. The TEI Header, v2.2.0, 25.08.2012, <<http://www.tei-c.org/Vault/P5/2.2.0/doc/tei-p5-doc/en/html/HD.html>> (30.08.2014); Dublin Core Metadata Element Set, v1.1, 14.06.2012 <<http://dublincore.org/documents/dces/>> (06.12.2012).
 - 5 TEI. P5: Guidelines for Electronic Text Encoding and Interchange. TEI element history, v2.2.0, 25.10.2012, <<http://www.tei-c.org/Vault/P5/2.2.0/doc/tei-p5-doc/en/html/ref-history.html>> (30.08.2014).
 - 6 Cf. Scholl et. al., Papyrus Projekt (see fn. 3 above), ch. 1.4 Datenmodell.

The metadata are collected in a FileMaker database ("fmp7" file format).⁷ For certain database fields, for example the "text topics keywords" field, we used "repeating fields".⁸ The contents of certain fields, for example place names, artefact types, and text types, are controlled by separate lookup tables.⁹ Rarely, the lookup tables actually restrict the possible values to certain possibilities (e.g. the values for the "constitution type" of the text as defined by the current TEI guidelines¹⁰). Usually, however, the lookup tables function as "thesauri" that normalize the set and spelling of certain values (e.g. "Thutmosis", not "Thutmose", "Thotmes", or else). For the initial content of these thesauri, we built upon data from data from the "Thesaurus Linguae Aegyptiae" (TLA), the "Berliner Papyrusdatenbank" (BerlPap) project, and the "Papyrus Projekt Leipzig".¹¹ Additionally, we seek to cross-reference texts, places, and ancient names to the respective thesauri of "Trismegistos".¹²

All data are generally encoded in Unicode compliant fonts. The main font used for the database was initially "Charis SIL". Since there is no free Unicode font available yet that contains the code ranges for all relevant scripts, the editors eventually have to change to specific Unicode fonts for the encoding of certain scripts, namely "New Athena Unicode" for Greek, "Antinoou" for Coptic, or "Aramaic Imperial Yeb" for Aramaic.¹³ For the transliteration of Pre-Coptic Egyptian we use the Unicode codes that I suggested elsewhere.¹⁴

7 FileMaker Pro v7–v11; see FileMaker Pro-Versionen im Vergleich, <<http://www.filemaker.de/products/filemaker-pro/version-comparison.html>> (30.08.2014).

8 The database design with "repeating fields" is convenient for the XSL Transformation to TEI XML, but it made special treatment necessary for the MySQL database import (see fn. 15 below).

9 The relations of the main metadata table and the thesauri were, however, not designed in a relational database fashion as preferred from an IT perspective. This was due to the fact that the philological editors needed to work with individual copies of the database, eventually without internet access available. A relational database architecture would potentially have made the subsequent process of merging the individual databases copies into one too time consuming and less robust.

10 TEI. P5: Guidelines for Electronic Text Encoding and Interchange. TEI element constitution, v2.2.0, 25.10.2012, <<http://www.tei-c.org/Vault/P5/2.2.0/doc/tei-p5-doc/en/html/ref-constitution.html>> (30.08.2014).

11 See fn. 3 above. Especially worth mentioning is that we use the lookup tables from the TLA that match Egyptian relative chronology with absolute dates. Although this table might not be up-to-date in every detail, the usage of it has the benefit that it facilitates a possible future import of relevant parts of our data into the TLA.

12 Trismegistos, <<http://www.trismegistos.org/>> (30.08.2014).

13 Fonts: Charis SIL, v4.110, <http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&id=charissilfont>; New Athena Unicode, v4.0.5, <<http://apagreekkeys.org/NAUdownload.html>>; Aramaic Imperial Yeb, v0.06-1, <<http://culmus.sourceforge.net/ancient/>>; Antinoou, v1.0.6, <<http://www.evertype.com/fonts/coptic/>> (all 30.08.2014). Keyboard layout for Greek and Coptic are available on the same websites. For a layout for Egyptological transliteration programmed by D.A.W., see Humboldt-Universität zu Berlin, Institut für Archäologie, LB Ägyptologie und Archäologie Nordostafrikas, Nützliches / Software & Links, <https://www.archaeologie.hu-berlin.de/aegy_anoa/utilities> (30.08.2014). For Imperial Aramaic a keyboard layouts was programmed using the program Microsoft Keyboard Layout Creator, v1.4, <<http://msdn.microsoft.com/de-de/goglobal/bb964665.aspx>> (12.12.2012).

14 Notably U+0131 & U+0357 for *l*, and the capital(sic) letter codes U+A722 and U+A724 for *ʒ* and *ç*, respectively. See Daniel A. Werning, Egyptological Transliteration Systems in Unicode, 17.01.2013, <http://www.archaeologie.hu-berlin.de/aegy_anoa/utilities/index_html/egyptological_transliteration_unicode> (30.08.2014); *idem* (user "dwer"), Table "Transkriptionssonderzeichen in Unicode", in: Wikipedia. Ägyptische Hieroglyphen, 04.01.2013, <http://de.wikipedia.org/w/index.php?title=Ägyptische_Hieroglyphen&oldid=112501095> (30.08.2014). For the erroneous rendering of the minuscules

The metadata will ultimately be exported from the FileMaker database in its proprietary “fmpdsresult” XML format and subsequently be transformed into two different XML formats. These are generated by specifically programmed eXtensible Stylesheet Language Transformations (XSL Transformations) operating on the “fmpdsresult” XML file. The first XML format to be generated is a ‘flat’¹⁵ XML format which serves as the source for the later MySQL import for the website, but also as a simple and easy to reuse primary archival format for the metadata. The second XSL Transformation generates a richly filled TEI XML header (<teiHeader>) from the metadata. For the suggested mapping of the collected metadata onto appropriate TEI header elements, see the appendix below. Compare the following (only slightly abbreviated) example:

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Pap. Berlin P. 13535</title>
      <editor>Lepper, Verena M.</editor>
      <respStmt><resp>Artefact metadata:</resp><persName>Moje, Jan</persName></respStmt>
      <respStmt><resp>TEI encoding:</resp><persName>Werning, Daniel A.</persName>
        </respStmt>
    </titleStmt>
    <publicationStmt><publisher> ... </publisher><pubPlace>Berlin</pubPlace></publicationStmt>
    <sourceDesc>
      <msDesc>
        <msIdentifier>
          <country>Deutschland</country>
          <settlement>Berlin</settlement>
          <institution>Staatliche Museen zu Berlin</institution>
          <repository>Ägyptisches Museum und Papyrussammlung</repository>
          <collection>Papyrussammlung</collection>
          <idno>Berlin P. 13535</idno>
          <msName>Pap. Berlin P. 13535</msName>
          <altIdentifier><idno>Berlin P. 535</idno></altIdentifier>
        </msIdentifier>
        <msContents>
          <summary>Vertrag als Königs-Eid: Verpflichtung von drei Personen zur gemeinsamen Eintreibung von verschiedenen Steuern. Mehrfachurkunde mit 1. Ausfertigung und 2 Zeilen der 2. Ausfertigung</summary>
          <msItem xml:id="RDB_TXT100297" defective="true">
            <locus>Recto</locus>
            <author/>
            <title/>
            <textLang mainLang="egy-demp-Egyd-Egydmd"/>
          </msItem>
        </msContents>
      </msDesc>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

of Egyptological Ain and Alef with a smaller height, see The Unicode Standard, Version 7.0. Latin Extended-D, 16.06.2014 <<http://www.unicode.org/charts/PDF/Unicode-7.0/U70-A720.pdf>> (30.08.2014).

- 15 The hierarchically organized structures of the “repeating fields” are partially combined and partially shifted to the basic hierarchical level to allow for a CSV transformation as input for the MySQL online database.

```
<objectDesc>
  <supportDesc xml:id="RDB_TR100297">
    <support>
      <dimensions>
        <height min="177" max="177" unit="mm"/>
        <width min="222" max="222" unit="mm"/>
        <depth>dünn</depth>
      </dimensions>
      <material style="color:papyrusfarben_dunkel">16 Papyrus</material>
      <objectType>Papyrus</objectType>
      <origDate/>
      <note>An der rechten Seite ist ein Schutzstreifen angeklebt. Der
        bei P. Eleph. Dem. 1, Nr. 11 erwähnte "scharfe Schnitt bis
        zur Hälfte der Blattkante" existiert jedoch nicht. </note>
    </support>
    <condition>
      <list>
        <item>Textträgerzustand:
          <measure>vollständig</measure>
          <desc>Unten abgeschnitten oder glatt abgebrochen,
            zwei Lücken in der letzten erhaltenen Zeile =Z.1
            der versehentlich nicht abgeschnittenen zweiten
            Ausfertigung der Schrift (Doppelurkunde).
            Horizontale Brüche entlang der Faltungskanten.
            Rechts eine vertikale Klebung (1 cm).</desc>
        </item>
        <item>Textzustand:
          <measure>unvollständig</measure>
          <desc>Unten abgebrochen. Pap. Berlin P.23677 passt
            unten links an.</desc>
        </item>
      </list>
      <join type="Joints" target="#RDB_TR100297 #RDB_TR100764">
        <desc>Pap. Berlin P. 13535 + Pap. Berlin P. 23677</desc>
      </join>
      <note>In Glasrahmen</note>
    </condition>
  </supportDesc>
  <layoutDesc><layout rendition="quer" columns="1" writtenLines="14">
    <height min="130" max="130" unit="mm" scope="columns"/>
    <width min="165" max="210" unit="mm" scope="rows"/>
    <note>Schriftspiegel Recto: 21 x 13 cm. Recto Erste Ausfertigung:
      13 Z. Recto Zweite Ausfertigung: 1+y Z. Eventuell ist
      ursprünglich mit drei Abschriften zu rechnen. </note>
  </layout></layoutDesc>
</objectDesc>
<handDesc><handNote scribeRef="#scribe"17> "Mitteldemotische
  Geschäftsschrift." (Nach Zauzich in VOHD XIX,2)</handNote>
</handDesc>
<scriptDesc><scriptNote medium="Tinte_schwarz_monochrom"/></scriptDesc>
```

16 As far as I see, there is no dedicated way to encode somewhat fuzzy information on material colors in the TEI header guidelines. The suggested code abuses the global Cascaded Style Sheet (CSS) attribute style as a preliminary workaround.

17 The encoding of information on scribes simply in the attribute @scribe of att.handFeatures is, regrettably, not possible since this attribute does only allow for single name strings.

```
</physDesc>
<history>
  <origin>
    <origDate>
      <date notBefore="-0244" notAfter="-0236">Juli 18 – Aug. 16 / Ptolemaios
        III., Reg.-Jahr 11, Pauni Alternative Datierung / Zauzich:
        Ptolemäisch, Reg.-Jahr 3</date>
      <certainty locus="value"><desc>quellensprachliche
        Datierung</desc></certainty>
    </origDate>
    <origPlace><country>Ägypten</country><region>Oberägypten, 1. Gau
      </region><settlement ref="http://www.trismegistos.org/place/621">
        Elephantine</settlement></origPlace>
  </origin>
  <provenance type="archäologische_Ausgrabung">
    <date notBefore="1906" notAfter="1906"/>
    <persName role="Finder">Rubensohn, Otto</persName>
    <location><country>Ägypten</country><region>Oberägypten, 1. Gau
      </region><settlement ref="http://www.trismegistos.org/place/621">
        Elephantine</settlement></location>
  </provenance>
  <acquisition>
    <date notBefore="1906" notAfter="1907"/>
    <location><country>Ägypten</country><region>Oberägypten, 1. Gau
      </region><settlement ref="http://www.trismegistos.org/place/621">
        Elephantine</settlement></location>
    <desc>Fundteilung; Blechkiste 227</desc>
    <note>12.02.1906</note>
  </acquisition>
</history>
<additional>
  <adminInfo>
    <availability>
      <licence>publiziert</licence>
      <ab><location/><desc>Depot | Papyrusdepot</desc></ab>
    </availability>
    <custodialHist><custEvent>Restaurierung abgeschlossen</custEvent>
      </custodialHist>
    <note>Zusammengehörig Pap. Berlin P. 13535 und Pap. Berlin P. 23677.
      1958 vorhanden</note>
  </adminInfo>
  <surrogates><list>
    <item><graphic url="P_13535+23677_R.jpg"/></item>
    <item><graphic url="P_13535+23677_V.jpg"/></item>
  </list></surrogates>
  <listBibl>
    <bibl>
      W. SPIEGELBERG, Demotische Papyrus von der Insel
      Elephantine I, Nr.1-13, Leipzig 1908 (Demotische Studien 2), Nr. 11,
      nur P. 13535
    </bibl>
    ...
    <bibl>
      <title>Trismegistos record</title>
      <idno>44287</idno>
```

```
<ref type="url">
  http://www.trismegistos.org/tm/detail.php?quick=44287</ref>
</bibl>
</listBibl>
</additional>
</msDesc>
<listPerson><personGrp xml:id="scribe">
  <persName role="Schreiber">Parates, Sohn des Payris (Zeile 1-12 Mitte)</persName>
  <persName role="Schreiber">Espmetis, Sohn des Djehuti-em-heb (Zeile 12 Mitte -
    13 Mitte)</persName>
  <persName role="Schreiber">Harpaesis, Sohn des Paweputi (Zeile 13 Mitte - 13
    Ende)</persName>
</personGrp></listPerson>
<listPerson type="im_Text_erwähnte"><personGrp>
  <persName>Parates, Sohn des Payris (Pa-rꜥ sꜥ Pa-hr)</persName>
  ...
  <persName>Ptolemaios III. ((Kartusche Pꜥtrwmys|Kartusche))</persName>
</personGrp></listPerson>
<listPlace type="im_Text_erwähnte">
  <place><placeName>Ägypten (Kmy)</placeName></place>
  <place><placeName>Bezirk von Elephantine (ḫꜥ shꜥ.t Yb)</placeName></place>
</listPlace>
</sourceDesc>
</fileDesc>
<encodingDesc><classDecl>
  <taxonomy xml:id="TAX_Textgattung"> ... <category xml:id="CAT_Vertrag"><catDesc>
    dokumentarisch | Vertrag</catDesc> </category> ... </taxonomy>
  <taxonomy xml:id="TAX_Sprachcodes"><category xml:id="egy-demp"><catDesc>Ägyptisch |
    Demotisch | ptolemäisches Demotisch</catDesc></category></taxonomy>
  <taxonomy xml:id="TAX_Schriftcodes"><category xml:id="Egyd-Egydmd"><catDesc>Ägyptisch |
    Demotisch | Mitteldemotisch</catDesc></category></taxonomy>
</classDecl></encodingDesc>
<profileDesc>
  <textDesc>
    <channel/>
    <constitution type="composite">fragmentierter Textträger, kompletter Text</constitution>
    <derivation/><domain/><factuality/><interaction/><preparedness/><purpose/>
  </textDesc>
  <langUsage><language ident="egy-demp-Egyd-Egydmd"/></langUsage>
  <textClass>
    <catRef target="#CAT_Vertrag" scheme="TAX_Textgattung"/>
    <keywords><term>Königseid</term><term>Salzsteuer</term><term>In-schen-Steuer</term>
    </keywords>
  </textClass>
</profileDesc>
<revisionDesc status="in_Arbeit"><listChange> ... </listChange></revisionDesc>
</teiHeader>
```

In a subsequent phase of the project, the texts that are inscribed on the artefacts will be encoded or transliterated, respectively, in Unicode (see above). And they will be organized and enriched with XML markups according to the TEI guidelines for the <text> body and, more specifically, along the lines of the more specific “EpiDoc

Guidelines" that are TEI compliant.¹⁸ Compare the following preliminary (slightly abbreviated) example encoding a Coptic ostracon:¹⁹

```
<text><body>
  <div type="inscription"><ab>
    <pb n="outside" facs="P_14803_R.jpg" />
    <lb n="1" rend="left-to-right" />ⲧ ⲁⲛⲐⲔ
    <lb n="2" /><persName type="private" key="Valentinus"
      ref="http://www.trismegistos.org/name/10883">ⲟⲩⲁⲗⲉⲧⲓⲛⲟⲥ</persName> <persName
      type="private" key="Johannes"
      ref="http://www.trismegistos.org/name/3464">ⲓⲟⲩⲁⲛⲛⲛⲉ</persName> ⲛⲣⲙⲉⲓⲛⲃ
    ...
    <lb n="4" />ⲛⲙⲁⲧⲁⲓ ⲛⲛⲁⲣⲓⲟⲓ<supplied reason="lost">ⲙ</supplied> ⲟⲥ ⲛⲉ<placeName type="settlement"
      key="Elephantine" ref="http://www.trismegistos.org/place/621">ⲉⲓⲛⲃ</placeName>
    ...
    <lb n="6" />ⲛⲁⲔ ⲩⲓⲥⲧⲉ ⲛⲕⲁⲣⲁ ⲛⲛⲟⲩⲃ ⲛⲁⲣ
    <pb n="inside" facs="P_14803_V.jpg" />
    <lb n="7" />ⲓ ⲣ ⲕ,ⲁ ⲉ ⲛⲁⲓ ⲧⲉⲛⲟⲩ ⲉⲓⲛⲟⲩⲧⲉ
    ...
    <lb n="11" />ⲉⲓⲛⲉⲣⲁ <date when-custom="Parmuti:9">ⲛⲁⲣ
    <lb n="12" />ⲙⲁⲧ ⲉ</date>
  </ab></div>
  <div type="translation" xml:lang="de"><ab>
    <pb n="outside" />
    <lb n="1" />Ich,
    <lb n="2" />Valentinus, (Sohn des) Johannes, Mann aus Elephantine:
    ...
    <lb n="4" />den Soldaten des Bataillons von Elephantine:
    ...
    <lb n="6" />dir neun Karat Gold im Nennwert,
    <pb n="inside" />
    <lb n="7" />macht 9 Karat. Diese nun – so Gott (es bestimmt) –
    ...
    <lb n="11" />Am heutigen Tage, Phar-
    <lb n="12" />muthi 9.
  </ab></div>
</body></text>
```

Note that the original text or transliteration, respectively, and the translation will visually appear on the webpage in a format according to the style that philologist of ancient languages are accustomed to, i.e. with markups following either the Leiden Conventions or Egyptological conventions.²⁰

18 See EpiDoc Guidelines, v8.16 <<http://www.stoa.org/epidoc/gl/8.16/>> (12.07.2014). Cf. also Papyri.info. Text Leiden+ Documentation, <papyri.info/editor/documentation?docotype=text> (30.08.2014).

19 Ostr. Berlin P. 14803; text and translation preliminarily encoded by Andrea Hasznos; TEI tags by D.A.W.

20 See Hans Krummrey, Silvio Panciera, Criteri di edizione e segni diacritici, in: *Miscellanea, Tituli 2* (1980), 205–215 (non vidi); Wikipedia, Leiden Conventions, 13.11.2012 <http://en.wikipedia.org/w/index.php?title=Leiden_Conventions&oldid=522779023> (27.02.2013); Wolfgang Schenkel, *Tübinger Einführung in die klassisch-ägyptische Sprache und Schrift*, 5th., [green] ed., Tübingen 2012, ch. 2.5.

§3 Research website

In order to give the user the possibility to comment on the provided information, we make use of the Open Source web blog Content Management System “WordPress”.²¹ The website of the Rubensohn Project will largely be modelled after the website of the Berliner Papyrusdatenbank (BerlPap) of the Staatliche Museen zu Berlin (SMB), which is supported by the Deutsche Forschungsgemeinschaft (DFG) since 2010.²² First of all, this is done in order to offer a similar user experience for the visitors who want to search in papyri and ostraca databases on the SMB server. Moreover, this may facilitate future cross-references and connections between the Berliner Papyrusdatenbank and the Rubensohn Project websites.

Besides the similarities in the visual design, the site map, and the users’ search experience, there are, however, also characteristic differences between the BerlPap website and the Rubensohn Project website. Since the text artefacts from Elephantine have not been systematically published yet, we don’t offer a dedicated search tree organized by primary editions as BerlPap has it. Besides a compact search form that includes only a dozen selected search fields, however, we also offer a (re)search form that encompasses nearly all of the more than 100 fields of the metadata database (including the bibliographies). Due to the fact that, differently to the case of the BerlPap corpus, all the artefacts come from the same place but are inscribed with a large variety different scripts, the information displayed in the intermediate search result list naturally encompass the scripts written on the objects rather than the place of origin.

The information page of a single artefact will offer the full set of information collected on the object. In addition to that, we aim at providing the philological encoding or transliteration, respectively, of the inscribed texts. We will also offer both, the metadata and the text, encoded in TEI XML format.

Finally, we are also going to provide a detailed documentation on the data IT of the Rubensohn project, i.e. the metadata, the TEI encoding strategies, as well as certain other created resources like the Unicode keyboard layouts for ancient scripts on the website.

§4 IT-related project workflow

The Digital Humanities (DH) project workflow was designed to balance the needs of the contributing philologists and the Information Technology (IT) tasks in a way that it is pragmatically efficient. Since the Rubensohn Project is to publish data and images of a limited set of ‘only’ approx. one thousand artefacts, some steps are actually more efficiently executed (semi-)manually despite the fact that more professional IT techniques could have been implemented.²³ **Fig. 1** displays the main parts of the

21 WordPress Deutschland, Versions-Archiv der DE-Editionen, v3.7.1, 29.10.2013 <<http://wpde.org/download/versions-archiv/>> (30.08.2014).

22 BerlPap. Berliner Papyrusdatenbank, 21.12.2012 <<http://www.smb.museum/berlpap/>> (30.08.2014). We would like to thank Dr. Fabian Reiter (SMB) and Marius Gerhardt (DFG) for sharing much of their internal data, and I am especially grateful to Paul Bartels (DFG) for further explanations of BerlPap’s IT model.

23 For example, given that he had to crop the objects images by hand, the responsible person chose to also paste the ruler and visual copyright information into the images by hand, simultaneously. We

workflow. Sustainable data is marked by bold borders. The grey parts are yet to be worked out.

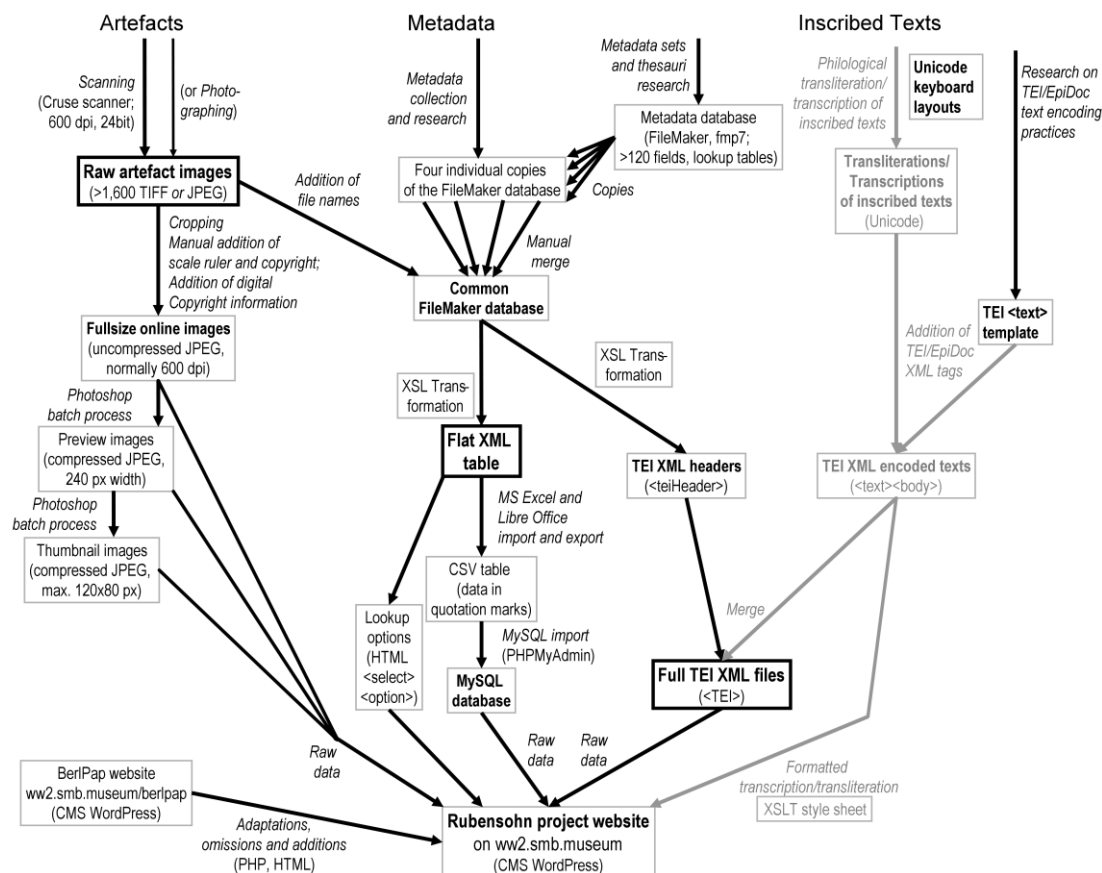


Fig. 1: IT-related workflow in the Rubensohn Project

Appendix: List of metadata fields and corresponding TEI header elements

No.	Information	Data type	TEI tags or attributes
Data record metadata			
1	Record ID	Txt, 6 digits	TEI[@xml:id]
2	Record name/title	Txt	fileDesc/titleStmt/title fileDesc/sourceDesc/msDesc/msIdentifier/msName
3a	Record editor	Txt, fixed	fileDesc/titleStmt/editor
3b	Record publisher, name	Txt, fixed	fileDesc/publicationStmt/publisher
3c	Record publisher, city	Txt, fixed	fileDesc/publicationStmt/pubPlace
4	Record authors	Txt	fileDesc/titleStmt/respStmt/resp"Artefact metadata:"/persName
5		Txt	fileDesc/titleStmt/respStmt/resp"TEI encoding:"/persName
6	Record creation date		revisionDesc/listChange/change[1]
7	Record change list	Txt[20x]	revisionDesc/listChange/change
8	Latest change of record	Date	revisionDesc/listChange/change[last()]
9	Record status	Txt, fixed set	revisionDesc[@status]

also did not struggle to set up a FileMaker Pro server but worked with separate copies of the metadata database during the phase of intensive distributed research, and later with a central copy stored on an online drive (thereby avoiding to deal with issues of eventually missing internet access and firewall policies on the SMB server).

No.	Information	Data type	TEI tags or attributes
10	Record, internal notes	Txt	
	Object/support metadata		fileDesc/sourceDesc/msDesc/...
11	Object/support ID	Txt	.../physDesc/objectDesc/supportDesc[@xml:id]
12a	Collection (modern), country	Txt, contr. set	.../msIdentifier/country
12b	Collection, city	Txt, contr. set	.../msIdentifier/settlement
12c	Collection, institution	Txt, contr. set	.../msIdentifier/institution
12d	Collection, museum	Txt, contr. set	.../msIdentifier/repository
13	Collection, inventory	Txt, contr. set	.../msIdentifier/collection
14	Collection/inventory (abbreviation)	Txt, contr. set	
15	Inventory no., primary, bare no.	Txt	
16	Inventory no., addition	Txt	
17	Inventory no., primary (full)	Txt	.../msIdentifier/idno
18	Alternative inventory nos.	Txt[3×]	.../msIdentifier/altIdentifier/idno
19	Fragmentation type	Txt, fixed set	profileDesc/textDesc/constitution[@type] profileDesc/textDesc/constitution (verbal)
20	Accessibility status, collection depot	Txt, contr. set	.../additional/adminInfo/availability/ab/desc
21	Collection depot/location of object	Txt	.../additional/adminInfo/availability/ab/location/note
22	Inventory/depot, notes	Txt	.../additional/adminInfo/availability/note
23	Publication permission status (public)	Txt, fixed set	.../additional/adminInfo/availability/licence
24	Provenance type	Txt, contr. set	.../history/provenance[@type]
25	Find spot, settlement	Txt, contr. set	.../history/provenance/location/settlement
26	Find spot, region	Txt, contr. set	.../history/provenance/location/region
27	Find spot, country	Txt, contr. set	.../history/provenance/location/country
28	Find spot, specific location	Txt	.../history/provenance/location/note
29a	Find date, year, <i>term. ante quem non</i>	Num	.../history/provenance/date[@notBefore]
29b	Find date, year, <i>term. post quem non</i>	Num	.../history/provenance/date[@notAfter]
30	Finder	Txt, contr. set	.../history/provenance/persName[@role="Finder"]
31	Antique archive	Txt	.../history/provenance/affiliation
32	Acquisition type	Txt, contr. set	.../history/acquisition/desc
33	Acquisition, city	Txt, contr. set	.../history/acquisition/location/settlement
34	Acquisition, region	Txt, contr. set	.../history/acquisition/location/region
35	Acquisition, country	Txt, contr. set	.../history/acquisition/location/country
36a	Acquisition date, year, <i>term. ante quem non</i>	Num	.../history/acquisition/date[@notBefore]
36b	Acquisition date, year, <i>term. post quem non</i>	Num	.../history/acquisition/date[@notAfter]
37	Acquisition, seller	Txt, contr. set	.../history/acquisition/listPerson/person/persName[@role="Verkäufer"]
38	Acquisition, buyer	Txt, contr. set	.../history/acquisition/listPerson/person/persName[@role="Käufer"]
39	Acquisition, shipping box	Txt	.../history/acquisition/desc
40	Acquisition, notes	Txt	.../history/acquisition/note
41	Material	Txt, contr. set	.../physDesc/objectDesc/supportDesc/support/material
42	Material, notes	Txt	.../physDesc/objectDesc/supportDesc/support/material/note
43	Text support color	Txt, contr. set	.../physDesc/objectDesc/supportDesc/support/material[@style="color: ___"] (see fn. 16 above)
44	Text support, dating (cf. data #98–100 below)	Txt	.../physDesc/objectDesc/supportDesc/support/date
45a	Text support, dating, year, <i>term. ante quem non</i>	Num	.../physDesc/objectDesc/supportDesc/support/origDate/ date[@notBefore]

<i>No.</i>	<i>Information</i>	<i>Data type</i>	<i>TEI tags or attributes</i>
45b	Text support, dating, year, <i>term. post quem non</i>	Num	.../physDesc/objectDesc/supportDesc/support/origDate/date[@notAfter]
46	Text support, dating criteria	Txt	.../physDesc/objectDesc/supportDesc/support/origDate/date/certainty/desc
47	Text support, dimensionality	Txt, fixed set	.../physDesc/objectDesc/supportDesc/support/dimensions/depth
48a	Text support, width, min. (mm)	Num	.../physDesc/objectDesc/supportDesc/support/dimensions/width[@min @unit="mm"]
48b	Text support, width, max. (mm)	Num	.../physDesc/objectDesc/supportDesc/support/dimensions/width[@max @unit="mm"]
49a	Text support, height, min. (mm)	Num	.../physDesc/objectDesc/supportDesc/support/dimensions/height[@min @unit="mm"]
49b	Text support, height, max. (mm)	Num	.../physDesc/objectDesc/supportDesc/support/dimensions/height[@max @unit="mm"]
50a	Text support, thickness or depth, min. (mm)	Num	.../physDesc/objectDesc/supportDesc/support/dimensions/depth[@min @unit="mm"]
50b	Text support, thickness or depth, max. (mm)	Num	.../physDesc/objectDesc/supportDesc/support/dimensions/depth[@max @unit="mm"]
51	Text support, preservation, approx. percentage	Txt, fixed set	.../physDesc/objectDesc/supportDesc/condition/list/item/"Textträgerzustand: "/measure
52	Text support, preservation, condition	Txt	.../physDesc/objectDesc/supportDesc/condition/list/item/desc
53	Text support, framing/mounting	Txt	.../physDesc/objectDesc/supportDesc/condition/note
54	Restoration(s)	Txt[3×]	.../additional/adminInfo/custodialHist/custodialEvent
55	Artefact type	Txt, contr. set	.../physDesc/objectDesc/supportDesc/support/objectType
56	Text support, further notes	Txt	.../physDesc/objectDesc/supportDesc/support/note
57	Text support, joints, database internal links	Txt[5×], ID links	.../physDesc/objectDesc/supportDesc/condition/join[@type="Joints" @target]
58	Text support, joints	Txt	.../physDesc/objectDesc/supportDesc/condition/join/desc
59	Object scans, file names	Txt[10×]	.../additional/surrogates/list/item/graphic[@url]
60	Photos in collection archive	Txt	.../additional/surrogates/list/item
61	Bibliography	Txt[20×]	.../additional/listBibl/bibl
62	Link to record on http://www.trismegistos.org	Num	.../additional/listBibl/bibl/title/"Trismegistos record" .../additional/listBibl/bibl/idno .../additional/listBibl/bibl/ref[@type="url"]
63	Link to record on http://ww2.smb.museum/berlpap	Txt	.../additional/listBibl/bibl/title/"BerlPap record" .../additional/listBibl/bibl/idno .../additional/listBibl/bibl/ref[@type="url"]
64	Link to record on http://papyri.info	Txt	.../additional/listBibl/bibl/title/"Papyri.info record" .../additional/listBibl/bibl/ref[@type="url"]
65	Further www links	Txt[5×]	.../additional/listBibl/bibl/ref[@type="url"]
Text metadata			fileDesc/sourceDesc/msDesc/...
66	Text ID	Txt	.../msContents/msItem[@xml:id]
67	Text location on support, side	Txt, contr. set	.../msContents/msItem/locus
68	Text location on support, detailed information	Txt	.../msContents/msItem/locus
69	Script, primary	Txt, contr. set	.../msContents/msItem/textLang[@mainLang] profileDesc/langUsage/language[@ident]
70	Language, primary	Txt, contr. set	.../msContents/msItem/textLang[@mainLang] profileDesc/langUsage/language[@ident]
71	Script, further A	Txt, contr. set	.../msContents/msItem/textLang[@otherLangs] profileDesc/langUsage/language[@ident]
72	Language, further A	Txt, contr. set	.../msContents/msItem/textLang[@otherLangs] profileDesc/langUsage/language[@ident]
73	Script, further B	Txt, contr. set	.../msContents/msItem/textLang[@otherLangs] profileDesc/langUsage/language[@ident]

No.	Information	Data type	TEI tags or attributes
74	Language, further B	Txt, contr. set	.../msContents/mslItem/textLang[@otherLangs] profileDesc/langUsage/language[@ident]
75	Scribal characteristics, notes	Txt	.../physDesc/handDesc/handNote
76	Script/fiber relation	Txt, fixed set	.../physDesc/objectDesc/layoutDesc/layout[@rendition]
77	Lines of text	Txt	.../physDesc/objectDesc/layoutDesc/layout[@writtenLines]
78	Columns of text	Txt	.../physDesc/objectDesc/layoutDesc/layout[@columns]
79a	Line lengths, min. (mm)	Num	.../physDesc/objectDesc/layoutDesc/layout/width[@scope="rows" @min @unit="mm"]
79b	Line lengths, max. (mm)	Num	.../physDesc/objectDesc/layoutDesc/layout/width[@scope="rows" @max @unit="mm"]
80a	Column heights, min. (mm)	Num	.../physDesc/objectDesc/layoutDesc/layout/height[@scope="columns" @min @unit="mm"]
80b	Column heights, max. (mm)	Num	.../physDesc/objectDesc/layoutDesc/layout/height[@scope="columns" @max @unit="mm"]
81	Text layout	Txt	.../physDesc/objectDesc/layoutDesc/layout/note
82	Script substance	Txt, contr. set	.../physDesc/scriptDesc/ScriptNote[@medium]
83	Text title, original ancient	Txt	.../msContents/mslItem/title[@type="antik"]
84	Text title, modern	Txt	.../msContents/mslItem/title[@type="modern"]
85	Text author, ancient (supposed)	Txt, contr. set	.../msContents/mslItem/author/persName
86	Scribe, ancient	Txt, contr. set	.../physDesc/handDesc/handNote[@scribeRef="#scribe] (see fn. 17 above) fileDesc/sourceDesc/listPerson/personGrp[@xml:id="scribe"/ persName[role="Schreiber"]
87	Preservation, defectiveness status (written text)	false, true	.../msContents/mslItem[@defective];
88	Preservation, approx. amount (written text; cf. data #51 above)	Txt, fixed set	.../physDesc/objectDesc/supportDesc/condition/list/item/"Textzustand: "/measure
89	Preservation, condition (written text; cf. data #52 above)	Txt	.../physDesc/objectDesc/supportDesc/condition/list/item/desc
90	Text class/category/type	Txt[3×], contr. set	profileDesc/textClass/catRef[@scheme="#TAX_Textgattung" @target] classDecl/taxonomy[@xml:id="TAX_Textgattung"/]category[@xml:id] classDecl/taxonomy[@xml:id="TAX_Textgattung"/]category/catDesc
91	Text subjects/topics, keywords	Txt[10×]	profileDesc/textClass/keywords/term
92	Names mentioned	Txt[20×]	fileDesc/sourceDesc/listPerson[type="im_Text_erwähnte"/]/personGrp/persName
93	Places mentioned	Txt[20×]	fileDesc/sourceDesc/listPlace[type="im_Text_erwähnte"/]/place/placeName
94	Text summary	Txt	.../msContents/summary
95	Place of creation (supposed), settlement	Txt, contr. set	.../history/origin/origPlace/settlement
96	Place of creation, region	Txt, contr. set	.../history/origin/origPlace/region
97	Place of creation, country	Txt, contr. set	.../history/origin/origPlace/country
98	Date of creation (text witness; cf. data #44/45)	Txt	.../history/origin/origDate/date
99a	Date of creation, year, <i>term. ante quem non</i>	Num	.../history/origin/origDate/date[@notBefore]
99b	Date of creation, year, <i>term. post quem non</i>	Num	.../history/origin/origDate/date[@notAfter]
100	Date of creation, dating criteria	Txt	.../history/origin/origDate/certainty/desc
101	Images/drawings accompanying text	Txt	.../physDesc/decoDesc/decoNote
102	Text, additional notes	Txt	.../msContents/mslItem/note

soll erscheinen in: Verena M. Lepper (Hrsg), [*Forschungen zur ägyptischen und orientalischen "Rubensohn-Bibliothek"*], Ägyptische und Orientalische Papyri und Handschriften des Ägyptischen Museums und Papyrussammlung Berlin, Berlin

x+14

<i>No.</i>	<i>Information</i>	<i>Data type</i>	<i>TEI tags or attributes</i>
103	Further texts on same text support, database internal link	Txt[5×], ID links	.../msContents/msItem/join[@type="andere Texte auf Textträger" @target]