
Daniel A. Werning

**TEI and Thesauri in the Rubensohn Project
(ÄMP Berlin)**

International Workshop

*Annotated Egyptian Corpora and TopBib Online – Exchange, Convergence,
Shared Objectives*

hosted by the Berlin-Brandenburgische Akademie der Wissenschaften

27–29 April 2015

The Rubensohn Project



Project Staff

Project Leader

Prof Dr. [Verena M. Lepper](#) (ÄMP)

E-mail: [v.lepper \(at\) smb.spk-berlin\(dot\)de](mailto:v.lepper@ smb.spk-berlin(dot)de)

- Homepage: <http://elephantine.smb.museum>
- Research database, primarily metadata
- >100 metadata fields

Scientific Staff

Dr. [Andrea Hasznos](#) (freie Mitarbeiterin, 2013–2014, koptische Textzeugen)

PD Dr. [Jan Moje](#) (ÄMP, 2012–2014, demotische Textzeugen)

Dr. [Matthias Müller](#) (freier Mitarbeiter, 2012–2014, hieratische Textzeugen)

Dr. [Daniel A. Werning](#) (freier Mitarbeiter, 2012–2015, Datenbank und Webauftritt)

PD Dr. [Stefan Wimmer](#) (freier Mitarbeiter, 2013–2014, aramäische Textzeugen)

IT Counsel and Advice

Team der *Berliner Papyrusdatenbank (BerlPap)*, insb. Paul Bartels

Dr. [Andreas Bienert](#) (SMB)

[Matthias Bruske](#) (SMB)

Metadata search forms

- Simple search form:

The screenshot shows a search interface for the Rubensohn Collection. At the top, there is a header with the title "Die ägyptische und orientalische Rubensohn-Bibliothek" and the logo of the "Ägyptisches Museum und Papyrussammlung, Staatliche Museen zu Berlin". Below the header is a navigation bar with links for "Suche", "Rubensohn-Sammlung", "Projekt", "Dokumentation & Hilfe", and "Kontakt". A search bar with a "Suchen" button and a flag icon is also present. The main search area contains several sections, each with a "Suchen" button, a "Felder leeren" button, and a "Zur Detailsuche" link. The sections include: "Inventarnummer" (text input), "Sammlungsstandort/-status" (dropdown), "Fundjahr: zwischen" (two date inputs with "und" in between), "Textträger" (dropdown), "Schriftarten" (dropdown), "Sprachen" (dropdown), "Datierung: zwischen (Jahr)" (two date inputs with "und" in between, with a note "(Negative Jahreszahlen, z.B. -632, entsprechen Jahren v.Chr.)"), "Textsorte/-gattung" (dropdown), "Textinhalt" (text input), "Bibliographie" (text input), "Trismegistos-Nr." (text input), and "Alle Felder" (text input). At the bottom, there is a note: "Kodezeichen "*" (Asterisk) ermöglicht Suche nach nicht leeren Feldern." and a footer with "Stand: 11. Apr. 2015 (v1.2) | Impressum & rechtliche Hinweise | Kontakt".

- Expert search form:
>100 metadata fields

The screenshot shows an expert search form with a large number of metadata fields. The form is organized into several sections, each with a "Suchen" button and a "Felder leeren" button. The fields include: "Inventarnummer", "Sammlungsstandort/-status", "Fundjahr: zwischen", "Textträger", "Schriftarten", "Sprachen", "Datierung: zwischen (Jahr)", "Textsorte/-gattung", "Textinhalt", "Bibliographie", "Trismegistos-Nr.", "Alle Felder", and many others. The form is designed to allow for highly specific searches across a wide range of metadata.

Search result

- Metadata HTML page (CC-BY-SA)
- Metadata download as TEI file (CC-BY-SA)
- Image of the manuscript (mostly CC-BY-NC-SA)
- *Next step*: Text, TEI/EpiDoc-encoded



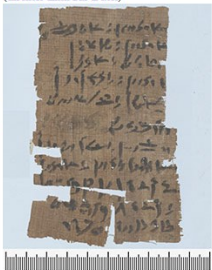
Pap. Berlin P. 15625

Persistent URL: <http://elephantine.smb.museum/record/ID100314/>

Datensatz-Zitation: Datensatz 100314 (= Pap. Berlin P. 15625), Projekt 'Die ägyptische und orientalische Rubensohn-Bibliothek', bearbeitet von J. Moje, hrsg. von Ägyptisches Museum und Papyrussammlung Berlin, <http://elephantine.smb.museum/record/ID100314/> (Stand 11.04.2015).

TEI-XML-Datei: http://elephantine.smb.museum/tei/elephantine_rubensohn_db_100314.tei.xml



Sammlung	Digitalbilder
Beherrbende Sammlung: Berlin, AMP, P.	P. 15625 Recto (direkter Link zur Datei)
Inventarnummer: Berlin P. 15625	
Sammlungstandort/-status: Depot Papyrusdepot:	
Fragmentierungstyp: einzelner, kompakter Textträger	
Publikationserlaubnis-Status: Publikationserlaubnis auf Nachfrage	
Herkunft	
Herkunftstyp: archäologische Ausgrabung	
Fundort: Elephantine, Oberägypten, 01. Gau, Ägypten :	
Fundjahr: zwischen 1906 und 1908	
Finder: Rubensohn, Otto / Zucker, Friedrich (Grabungleiter)	
Erwerbungsstyp: Fundteilung	
Erwerbungsart: Elephantine Oberägypten, 01. Gau, Ägypten	
Erwerbungsart: zwischen 1906 und 1908	
Textträger	P. 15625 Verso (direkter Link zur Datei)

This is a detailed view of the metadata for 'Pap. Berlin P. 15625'. It includes sections for 'Sammlung' (Collection), 'Titel' (Title), 'Beschreibung' (Description), 'Textträger' (Text carrier), and 'Anmerkungen' (Remarks). The 'Beschreibung' section contains a large amount of text, including the title 'P. 15625 Recto' and 'P. 15625 Verso'. The 'Textträger' section includes the title 'P. 15625 Recto' and 'P. 15625 Verso'. The 'Anmerkungen' section contains a large amount of text, including the title 'P. 15625 Recto' and 'P. 15625 Verso'. There are also thumbnail images of the papyrus fragments on the right side of the page.

Rich metadata in <teiHeader>

- More than 100 pieces of information.
- Best match with TEI element meanings (no reinterpretation)
=> **TEI P5 All.**
- TEI file creation:
 - XSLTransformation of a Filemaker database export.

BTW:

- Difficult to encode:
 - Text support color.

```
<layoutDesc>
  <layout writtenLines="14" />
</layoutDesc>
</objectDesc>
</physDesc>
<history>
  <origin>
    <origDate>
      <date notBefore="-236" notAfter="-236">BCE 236 Juli 18 - Aug. 16 (Ptolemaios III., Reg.-Jahr
      alternative Datierung (Zauzich): Ptolemäisch, Reg.-Jahr 03)</date>
    </origDate>
    <origPlace>
      <country>Ägypten</country>
      <region>Oberägypten, 01. Gau</region>
      <settlement>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, 01. Gau</region>
      <settlement>Elephantine</settlement>
    </location>
  </provenance>
  <acquisition>
    <note>Aus der Ausgrabung des Hr. Dr. O. Rubensohn in Elephantine 1906 (Bl. 227)</note>
  </acquisition>
</history>
<additional>
  <adminInfo>
    <note>zusammengehörig P. 13535 und P. 23677; 1958 vorhanden</note>
  </adminInfo>
  <listBibl>
    <bibl>P. Eleph. Dem. 1, Nr. 11 P. Berl. Eleph. 2 (= DPB III, Nr. 13535 & 23677 </bibl>
    <bibl>P. L. Bat. 9, 220 </bibl>
    <bibl>P. Äg. Handschr. 2 (= VOHD 19,2), Nr. 9 und 304 (Kurzbeschreibung) </bibl>
    <bibl>P. Eleph. Engl., 363-365 Nr. C 32 DBL (2005), 13, 51 - 52, 203</bibl>
    <bibl>
      <ref type="url">http://www.trismegistos.org/tm/detail.php?quick=44287</ref>
```

XSLTransformation

- Filemaker database

- Mapping Filemaker XML => TEI P5 All XML with help of *Altova MapForce*

The image shows a complex mapping process. On the left, the FileMaker Pro interface displays a database record for 'Pap. Berlin P. 13528'. The central area is the Altova MapForce mapping editor, showing a dense network of data flow lines connecting FileMaker fields to XSLT transformation rules. On the right, the resulting TEI XML file is shown, containing detailed metadata for the document, including its date (1906), location (Elephantine), and acquisition details.

• TEI file

<teiHeader>-specific issues

- Language and script encoding by tags conforming to [RFC 5646](#), [BCP 47](#), e.g. “egy-Egyh”
- Demand for standardization for **chronolects** and **subtypes of scripts**, e.g. Middle Egyptian, Late Egyptian, ..., or: Late Hieratic, ..., Bohairic
- Examples for current Rubensohn Project tags:
 - Middle Egyptian in classical Hieroglyphs: „egy-egym-Egyp-Egypreg“
 - Bohairic Coptic: „cop-copb-Copt“

zurück zu Hauptdatenbank			
deutsch	engl	scriptCode	+
Ägyptisch Hieratisch Archaisches Hieratisch		Egyn-Egynoi	
Ägyptisch Hieratisch Hieratisch		Egyh-Egyhar	
Ägyptisch Hieratisch Mittelhieratisch		Egyh-Egyhmd	
Ägyptisch Hieratisch Mittelhieratisch Mittelhieratische		Egyh-Egyhmdb	
Ägyptisch Hieratisch Mittelhieratisch Mittelhieratische		Egyh-Egyhmdb	
Ägyptisch Hieratisch Neuhieratisch		Egyh-Egyhlt	
Ägyptisch Hieratisch Neuhieratisch Neuhieratische Buchschrift		Egyh-Egyhltb	
Ägyptisch Hieratisch Neuhieratisch Neuhieratische Kanzleischrift		Egyh-Egyhltc	
Ägyptisch Hieratisch Späthieratisch		Egyh-Egyhlr	
Ägyptisch Hieratisch Späthieratisch Späthieratische Buchschrift		Egyh-Egyhlrb	
▶ Ägyptisch Hieroglyphen	Egyptian	Egyp	
Ägyptisch Hieroglyphen kryptographische Verwendung		Egyp-Egypcr	
Ägyptisch Hieroglyphen Ptolemäisch		Egyp-Egyppt	
Ägyptisch Hieroglyphen, klassische	Egyptian	Egyp-Egypreg	
Ägyptisch Kursivhieroglyphen	Egyptian	Egyp-Egypcs	
Arabisch	Arabic	Arab	
Aramäisch, Imperial	Aramäisch,	Armi	
Griechisch	Greek	Grek	
Koptisch	Coptic	Copt	
Koptisch Koptische Halbunziale		Copt-Copthu	
Koptisch Koptische Kursive		Copt-Coptcs	
Koptisch Koptische Unziale		Copt-Coptu	
Lateinisch	Latin	Latn	
Meroitische Hieroglyphen	Meroitic	Mero	
Meroitische Kursive	Meroitic	Merc	
(nicht identifiziert)		Unidentified	

Thesauri: demand of/usage in Rubensohn Project

Thesaurus	Source
Languages	<i>own list</i> , inspired by TLA list
Language tags (e.g. <i>egy</i> , <i>egym</i>)	IANA registry plus many own additions
Scripts	<i>own list</i> , inspired by TLA list
Script tags (e.g. <i>Egyh</i>)	IANA registry plus many own additions
Place names	Trismegistos GEO no.
Personal names Encoding: <persName type="private" key="Valentinus" ref=" http://www.trismegistos.org/name/10883 " >ΟΥΑΛΕΤΙΝΟC</persName>	(Rubensohn Project no. and 'standardized' name spelling) Trismegistos name no. (if existent) Trismegistos person no. (if existent)
Regnal years => absolute dates Encoding: <date notBefore="-236" notAfter="-236"> <i>Ptolemaios III., Reg.-Jahr 11, Pauni</i> </date>	Table from TLA

Thesauri: demand of/usage in Rubensohn Project

Thesaurus	Source
Text types/genres (e.g. <i>documentary</i> <i>name list</i>)	<i>own compilation</i> , inspired by Papyrus-Projekt Halle—Leipzig—Jena, Papyrus Portal, Berliner Papyrusdatenbank, TLA
Text support type (e.g. <i>ostrakon</i> , <i>papyrus</i>)	
Text support material (e.g. <i>stone</i> <i>sandstone</i>)	
Text support color (e.g. <i>pottery color</i> <i>light brown</i>)	<i>own compilation</i> , inspired by Papyrus-Projekt Halle—Leipzig—Jena
Text position (e.g. <i>recto</i> , <i>recto/verso</i> , <i>flesh side</i>)	Berliner Papyrusdatenbank <i>plus own additions</i>
Inscription substance (e.g. <i>ink</i> <i>bichrome</i>)	<i>own compilation</i>

See https://wikis.hu-berlin.de/annotated_text_databases/Main_Page#Metadata_Thesauri

TEI <text> – blueprint

```
<text>
<body>
  <div type="inscription">
    <ab>
      <pb n="outside" facs="http://elephantine.smb.museum/bilder/fullsize/
        /P_14803_R.tif" />
      <lb n="1" rend="left-to-right" />† ΔΝΟΚ
      <lb n="2" /><persName type="private" key="Valentinus"
        ref="http://www.trismegistos.org/name/10883">ΟΥΔΔΣΤΙΝΟΟ</persName>
      <persName type="private" key="Johannes"
        ref="http://www.trismegistos.org/name/3464">ΙΩΔΑΝΝΙΣ</persName>
      ΠΡΜΘΙΗΒ
      <lb n="3" />ΘΥΟΑΔ Ν<persName type="private"
        key="Paham">ΠΔΔΑΜ</persName> <persName type="private"
        key="Abraham">ΔΒΡΑΜ</persName>
      <lb n="4" />ΠΜΔΤΑΙ ΝΠΔΡΙΘΙ<supplied reason="lost"
        cert="high">Μ</supplied>ΟΟ Ν<placeName type="settlement"
        key="Elephantine"
        ref="http://www.trismegistos.org/place/621">ΘΙΗΒ</placeName>
      <lb n="5" />ΧΘ †ΖΟΜΟΛΙΓΙ ΔΥΩ †ΧΡΩΟΤΙ
      <lb n="6" />ΝΔΚ ΨΙΟΤΕ ΝΚΑΡΑ ΝΝΟΥΒ ΝΔΡ
      <pb n="inside" facs="http://elephantine.smb.museum/bilder/fullsize/
        P_14803_V.tif" />
      <lb n="7" />Γ ΡΚ,Δ Θ ΝΔΙ ΤΕΝΟΥ ΘΠΝΟΥΤΕ
      <lb n="8" />ΤΔΔΥ ΝΔΚ ΝΔΥ ΝΙΜ ΘΙΚΑΝ ΨΝΘ
      <lb n="9" />ΝΟΟΟΥ ΤΔΔΥ ΝΔΚ ΔΣΠ ΛΔΔΥ
      <lb n="10" />ΝΖΩΒ ΝΨΔΧΘ
      <lb n="11" />ΖΗΜΕΡΑ <date when-custom="Parmuti:9">ΠΔΡ
      <lb n="12" />ΜΔΤ Θ</date>
    </ab>
  </div>
```

- No stand-off markup (necessary).

Rubensohn Project encoding list (largely EpiDoc)

Description	Trad. philolog. markup	TEI/EpiDoc markup (cf. http://www.stoa.org/epidoc/gl/latest/app-alltrans.html)	Reference
Modern reconstruction in gap based on parallel text witnesses	s[Ptwlmys] irm	s[<supplied reason = "lost" evidence = "parallel" > Ptwlmys </supplied >] irm	http://www.stoa.org/epidoc/gl/latest/trans-parallelcopy.html
Characters now lost, but recorded by earlier scholars	j[mn .t	j <supplied reason = "lost" evidence = "previous_editor" > mn </supplied > .t	http://www.stoa.org/epidoc/gl/latest/trans-previouseditor.html
Partial destruction, but perfectly or unambiguously legible	jmn .t	jmn <damage > .t </damage > jmn <damage degree = "medium" > .t </damage > jmn <damage degree = "low" > .t </damage >	http://www.stoa.org/epidoc/gl/latest/trans-damaged.html https://wikis.huberlin.de/annotated_text_databases/TEI/Body
Partial destruction, interpretation somewhat ambiguous	jmn .t δεδα α τα	jmn <unclear reason = "damaged" > .t </unclear > δεδα <unclear reason = "damaged" > α </unclear > τα	http://www.stoa.org/epidoc/gl/latest/trans-ambiguous.html
MODERN EDITORIAL CORRECTIONS, EMENDATIONS, IDEALIZATIONS, ...			
Erroneous ancient addition	jmn .t }r ^c	jmn <surplus reason = "mistake" > t </surplus > -r ^c jmn <surplus reason = "mistake" cert = "high" > t </surplus > -r ^c jmn <surplus reason = "mistake" cert = "low" > t </surplus > -r ^c	https://wikis.huberlin.de/annotated_text_databases/TEI/Body http://www.stoa.org/epidoc/gl/latest/trans-superfluous.html
Erroneous ancient repetition	j{j}mn.t {jmn} jmn.t	j <surplus reason = "repeated" > j </surplus > mn.t j <surplus reason = "repeated" cert = "high" > j </surplus > mn.t j <surplus reason = "repeated" cert = "low" > j </surplus > mn.t	http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-surplus.html
Traditional 'superfluous' double spelling	zw{r}j t{r}j	zw <surplus reason = "traditional_orthography" > r </surplus > j	https://wikis.huberlin.de/annotated_text_databases/TEI/Body
Erroneous ancient omission	jmn .w >	jmn <supplied reason = "omitted" > .w </supplied > imn <supplied reason = "omitted" cert = "middle" > .w </supplied >	http://www.stoa.org/epidoc/gl/latest/trans-omitted.html

Links

- **Rubensohn-Projektseite:** <http://elephantine.smb.museum>
Dokumentation: <http://elephantine.smb.museum/dokumentation/>
- Daniel A. Werning. **Information Technology and Digital Humanities Workflow in the Rubensohn Project: Research Website and Rich TEI XML Header Encoding**, 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, approx. 14 pages, http://wwwuser.gwdg.de/~dwernin/drafts/Werning-Rubensohn_Projekt_IT-Manuskript.pdf.
- Daniel A. Werning. Sept. 2013. **Rubensohn-Datenbank: Datenfelder der Haupttabelle und TEI-Tag-Zuordnung**, http://elephantine.smb.museum/wp-content/uploads/Werning-RubensohnDB-Felder_Haupttabelle-Sept2013.pdf.
- **Annotated Text Databases. A collaborative Wiki for the coordination of TEI encoding and metadata thesauri for ancient manuscripts** (Open access Wiki) http://wikis.hu-berlin.de/annotated_text_databases/, ed. by Daniel A. Werning, Berlin: Humboldt University Berlin.
- **Glossing Ancient Languages** (Open access Wiki) http://wikis.hu-berlin.de/interlinear_glossing/, ed. by Daniel A. Werning, Berlin: Humboldt University Berlin.