Coptic Studies on the Digital Frontier: Creative Approaches to Manuscript Publication

Caroline T. Schroeder
University of the Pacific
www.carrieschroeder.com

November 17, 2012
Society of Biblical Literature Annual Meeting:
Manuscripts from Eastern Christian Traditions
I. Why We Need Digital Editions and Digital Coptic Resources

II. Technologies Available for Digital Editions

III. Databases and Computer-aided Research
I. WHY WE NEED DIGITAL EDITIONS AND DIGITAL COPTIC RESOURCES
Coptic literary texts

Basic access to sources

Computational/computer-aided research

*Digital humanities is not just a fad*
Access & Shenoute’s Corpus

Most important corpus for Sahidic Coptic

One of most important corpora for early monasticism (not hagiographical)

Image credits: Bib. Nat. Paris
Access & Shenoute’s Corpus

- Sermons
- Letters
- Treatises
- Monastic rules

Image credits: Bib. Nat. Paris
Access & Shenoute’s Corpus

- Sermons
- Letters
- Treatises
- Monastic rules
- Folios scattered
- Texts fragmented
- No crit eds
- Few publications
- Fewer translations

Image credits: Bib. Nat. Paris
Critical editions in preparation since 2000 by team of senior Coptic scholars; timetable uncertain

In the interim: Shenoute scholarship has slowed

What Ph.D. student will take on an edition as a dissertation knowing these are forthcoming?

What print publisher will publish someone else’s edition knowing these are forthcoming?
Computational Research

Gospel of Jesus’s Wife

Scholarly debates over...

- **Word frequency**: how likely is it that such hot-button vocabulary ("wife," "Jesus," "disciple," "dwell") would appear in such a small fragment?

C. Schroeder, SBL 2012
Persuasive, but...

...wouldn’t a larger corpus in which you could search for word frequencies, patterns of copying in authenticated ancient texts, etc., be more persuasive...?

Computational Research

Gospel of Jesus’s Wife

Scholarly debates over...

• **Word frequency**: how likely is it that such hot-button vocabulary ("wife," "Jesus," "disciple," "dwell") would appear in such a small fragment?

• **Grammar and Syntax**: is a missing ⲙ in the papyrus copied from a typo in a modern online G. Th. Edition?

C. Schroeder, SBL 2012
Again persuasive…

…but again, wouldn’t it be more persuasive if you could search for these phenomena in actual ancient Coptic manuscripts?

Evidence of copying a modern typo?

accessibility
AND
computer-aided research
primary desiderata
II. TECHNOLOGIES AVAILABLE FOR DIGITAL EDITIONS
First surviving folio of first letter of Shenoute

- Before Shenoute was leader of the monastery
- List of rules
  - By a predecessor?
  - Shenoute edited his own corpus → written or edited by him?

C. Schroeder, SBL 2012
Digital Edition Can Provide...

1. Manuscript photographs
2. Diplomatic edition
3. Normalized edition
4. If multiple manuscripts, linked editions
5. Links to related or re-used texts (biblical quotations, other textual citations and allusions, other monastic rules...)
6. Linked data: geographic data, online encyclopedias, etc.
7. Encoded data for current and future research
8. Translations
Digital Edition Can Provide...

1. Manuscript photographs
2. Diplomatic edition
3. Normalized edition
4. If multiple manuscripts, linked editions
5. Links to related or re-used texts (biblical quotations, other textual citations and allusions, other monastic rules...)
6. Linked data: geographic data, online encyclopedias, etc.
7. Encoded data for current and future research
8. Translations
White Monastery Manuscript XC 7-8, K9101 at the Österreichische Nationalbibliothek

http://www.earlymonasticism.org -- click on Pcol’s Rule
EpiDoc encoding:
• Papyri & epigraphy
• Leiden Conventions for manuscript markup
• Becoming standard in field
• Compliant with Text Encoding Initiative (TEI) XML standards
• EpiDoc provides XML schemas and XSL transformations

http://epidoc.sourceforge.net/
http://www.tei-c.org
EpiDoc TEI Markup: Encoded data according to recognized standards

1. Manuscript photographs
2. Diplomatic edition
3. Normalized edition
4. If multiple manuscripts, linked editions
5. Links to related or re-used texts (biblical quotations, other textual citations and allusions, other monastic rules...)
6. Linked data: geographic data, online encyclopedias, etc.
7. Encoded data for current and future research
8. Translations

XML can be imported into a database for search
XML can be visualized on webpage with EpiDoc transformations
EpiDoc TEI Markup:
Encoded data according to recognized standards

- XML can be imported into a database for search
- XML can be visualized on webpage with EpiDoc transformations

Not entirely “out-of-the box”
- Requires knowledge of XML and HTML
- Encoding and visualization does not entirely work for manuscript marks in Coptic literary texts
EpiDoc TEI Markup: Encoded data according to recognized standards

1. Manuscript photographs
2. Diplomatic edition
3. Normalized edition
4. If multiple manuscripts, linked editions
5. Links to related or re-used texts (biblical quotations, other textual citations and allusions, other monastic rules...)
6. Linked data: geographic data, online encyclopedias, etc.
7. Encoded data for current and future research
8. Translations

C. Schroeder, SBL 2012
Out-of-the-box Manuscript Transcription?

T-Pen

http://t-pen.org

• Upload manuscript image
Out-of-the-box Manuscript Transcription?

- Upload manuscript image
- Automatically breaks it into lines and columns (can be manually adjusted)
- Transcribe text
Out-of-the-box Manuscript Transcription?

- Upload manuscript image
- Automatically breaks it into lines and columns (can be manually adjusted)
- Transcribe text
- Export HTML or XML file
Out-of-the-box Manuscript Transcription?

T-Pen generated HTML file

Didn’t quite work out of the box...

No project indicated
T-Pen generated HTML file

Vienna, ONB Papyrussammlung K9101

Group Members:

- Schroeder (carrie@carrieschroeder.com) *Group Leader*
- Patrick Cuba (cubap@slu.edu)

Project Metadata

| Title: Vienna, ONB Papyrussammlung K9101 | Subtitle: | MS identifier: WM XC 7-8 |
| Subject: | MS Repository: ONB | MS Settlemnet: |
| Language: Coptic | Author: Shenoute | MS Collection: Papyrussammlung |
| Description: | Date: | MS id number: K9101 |

K-9101-RV_1.jpg

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Δ1</td>
<td>ὑγειευματαρ</td>
<td></td>
</tr>
<tr>
<td>Δ2</td>
<td>ἀθεοβολεχ</td>
<td></td>
</tr>
<tr>
<td>Δ3</td>
<td>ὀγχοοττεματ</td>
<td></td>
</tr>
<tr>
<td>Δ4</td>
<td>καταλαγενη</td>
<td></td>
</tr>
<tr>
<td>Δ5</td>
<td>Φεόγορησιανο</td>
<td></td>
</tr>
<tr>
<td>Δ6</td>
<td>διογκεχαλόμ</td>
<td></td>
</tr>
<tr>
<td>Δ7</td>
<td>σιαλερομε</td>
<td></td>
</tr>
</tbody>
</table>
Almost out-of-the-box Manuscript Transcription?

Hiccups:
Still don’t know the cause, but creators are working on it
Problems with uploading transcription text, as well
Possibly content (Coptic Unicode) related?

Advantages:
Great user-support via email
Ideal for collaboration
Can use TEI-compliant XML tags OR user-defined tags
Somewhat “out-of-the-box”
Translation: Alpheios Tool

http://alpheios.net.

Not visible here: hovering over the text highlights text and corresponding translation
III. COMPUTATIONAL AND COMPUTER-AIDED RESEARCH
Coptic SCRIPItoriuM

(Sahidic Corpus Research: Internet Platform for Interdisciplinary multilayer Methods)
Coptic SCRIPtorluM

Collaboration

Caroline T. Schroeder (University of the Pacific, Religious and Classical Studies)

Amir Zeldes (Humboldt University, Institute for German Language and Linguistics)
Coptic SCRIPtorIuM

1. Minimal TEI-XML editions of texts (no photographs)

2. Software tools for processing and annotating Coptic texts

3. A searchable database of Coptic for historical and linguistic research

Free and open source

C. Schroeder, SBL 2012
1. Minimal TEI-XML editions of texts (no photographs)
2. Software tools for processing and annotating Coptic texts
3. A searchable database of Coptic for historical and linguistic research

*Free and open source*
Under the hood: XC 7-8
(-using Exmaralda)

Multiple layers of information in the database, beginning with the text
Under the hood: XC 7-8
(using Exmaralda)

Diplomatic edition: column breaks
Under the hood: XC 7-8
(using Exmaralda)

Diplomatic edition: line breaks
Under the hood: XC 7-8
(using Exmaralda)

Diplomatic edition: typical word construction
Under the hood: XC 7-8
(using Exmaralda)

Diplomatic edition: typical word construction
Coptic SCRIPtorIuM

1. Minimal TEI-XML editions of texts (no photographs)
2. Software tools for processing and annotating Coptic texts
3. A searchable database of Coptic for historical and linguistic research

*Free and open source*
Tools to automate or semi-automate:

• Tokenizing AND...
• Lemmatizing
• Part-of-speech-tagging
Potential research questions computational research can investigate:

• Vocabulary studies
  – Historical implications: evolution of the concept of the “novice” monk; usage of theological terms; etc.

• Grammar, syntax, and other linguistic questions

• Writing style

• Textual re-use (quotations, citations, allusions)

• Bilingualism
  – Can a study of vocabulary, style, syntax, and grammar help us determine whether Shenoute was bilingual?
Potential research questions computational research can investigate in XC 7-8 (long-term):

• Was Shenoute the author? (vocabulary, style)
• Where else do these rules get quoted, cited, referenced – in Shenoute, Pachomius, etc.?
• Are there *patterns* to the ways or contexts in which the rules are quoted or referenced
Stay tuned.
www.earlymonasticism.org

With thanks to...
- The National Endowment for the Humanities
- The Perseus Digital Library (Greg Crane, Bridget Almas, et al.)
- Amir Zeldes
- Anke Lüdeling
- The Österreichische Nationalbibliothek
- David J. Birnbaum
- Frederik Krueger