Digital tools for scholarly editions

An overview, and a wish list
Humanities in the digital age...
...so what is a digital edition?
And how do we make one?
There are a lot of tools to cover, rather than subjecting you to a laundry list I’m going to give you this URL, and spend the time talking about the shape of what’s out there.

You can think of the process of critical edition in several stages, most of which have some useful resources and/or tools...
Tools for critical edition

...or you can think of the resources and tools according to a few categories. [TRANS for each column]

I’ll give an overview of the sorts of things we have in each of the categories, and where the gaps are, particularly when it comes to Oriental languages
– Don’t have an exhaustive sense of resources, so Coptic/Arabic/Syriac specialists invited to add what they know, to go on the wiki
Tools for critical edition

Information gathering

- Discovery of current project(s)
- Manuscript search
- Manuscript acquisition

...or you can think of the resources and tools according to a few categories. [TRANS for each column]

I’ll give an overview of the sorts of things we have in each of the categories, and where the gaps are, particularly when it comes to Oriental languages

- Don’t have an exhaustive sense of resources, so Coptic/Arabic/Syriac specialists invited to add what they know, to go on the wiki
## Tools for critical edition

<table>
<thead>
<tr>
<th>Information gathering</th>
<th>Data creation</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Discovery of current project(s)</td>
<td>• Scanning/OCR</td>
</tr>
<tr>
<td>• Manuscript search</td>
<td>• Transcription</td>
</tr>
<tr>
<td>• Manuscript acquisition</td>
<td>• Collation</td>
</tr>
</tbody>
</table>

...or you can think of the resources and tools according to a few categories. [TRANS for each column]

I’ll give an overview of the sorts of things we have in each of the categories, and where the gaps are, particularly when it comes to Oriental languages

– Don’t have an exhaustive sense of resources, so Coptic/Arabic/Syriac specialists invited to add what they know, to go on the wiki
### Tools for critical edition

<table>
<thead>
<tr>
<th>Information gathering</th>
<th>Data creation</th>
<th>Data analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Discovery of current project(s)</td>
<td>• Scanning/OCR</td>
<td>• Stemmatic analysis</td>
</tr>
<tr>
<td>• Manuscript search</td>
<td>• Transcription</td>
<td>• Criticism of the texts and data</td>
</tr>
<tr>
<td>• Manuscript acquisition</td>
<td>• Collation</td>
<td></td>
</tr>
</tbody>
</table>

...or you can think of the resources and tools according to a few categories. [TRANS for each column]

I’ll give an overview of the sorts of things we have in each of the categories, and where the gaps are, particularly when it comes to Oriental languages

– Don’t have an exhaustive sense of resources, so Coptic/Arabic/Syriac specialists invited to add what they know, to go on the wiki
...or you can think of the resources and tools according to a few categories. [TRANS for each column]

I’ll give an overview of the sorts of things we have in each of the categories, and where the gaps are, particularly when it comes to Oriental languages

– Don’t have an exhaustive sense of resources, so Coptic/Arabic/Syriac specialists invited to add what they know, to go on the wiki

<table>
<thead>
<tr>
<th>Information gathering</th>
<th>Data creation</th>
<th>Data analysis</th>
<th>Data publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovery of current project(s)</td>
<td>Scanning/OCR</td>
<td>Stemmatic analysis</td>
<td>Formatting and publication</td>
</tr>
<tr>
<td>Manuscript search</td>
<td>Transcription</td>
<td>Criticism of the texts and data</td>
<td></td>
</tr>
<tr>
<td>Manuscript acquisition</td>
<td>Collation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tara L. Andrews | Katholieke Universiteit Leuven | 26 October 2010
Information gathering

- You want to know if someone is working on this text that needs a new edition.
- You have your text and want to find the MSS.
- Catalogues of collections; most printed, some (e.g. obscure Armenian collections) v. hard to find; a very few digitized.
- Getting copies; mostly through individual libraries, often not online for us, but starting to get better. Notably Hill.
Information gathering

- You want to know if someone is working on this text that needs a new edition.
- You have your text and want to find the MSS.
- Catalogues of collections; most printed, some (e.g. obscure Armenian collections) v. hard to find; a very few digitized.
- Getting copies; mostly through individual libraries, often not online for us, but starting to get better. Notably Hill.
Info gathering.
- You want to know if someone is working on this text that needs a new edition.
- You have your text and want to find the MSS.
- Catalogues of collections; most printed, some (e.g. obscure Armenian collections) v. hard to find; a very few digitized.
- Getting copies; mostly through individual libraries, often not online for us, but starting to get better. Notably Hill.
Data creation

- OCR tools
- Transcription tools
- Collation tools (manual collation)

Now have collected the MSS, time to start turning it into text. First step is to get the text from parchment to screen:
- OCR: some help, sometimes
- Transcription: gotta be done sometime
- Collation: if you are doing it manually
If you’re doing collation by hand, this is where you find the tools most of you are familiar with. Spreadsheets, and CTE. Not going to dwell on these; CTE to be discussed further on in this session.
If you’re doing collation by hand, this is where you find the tools most of you are familiar with. Spreadsheets, and CTE. Not going to dwell on these; CTE to be discussed further on in this session.
If you’re doing collation by hand, this is where you find the tools most of you are familiar with. Spreadsheets, and CTE.

Not going to dwell on these; CTE to be discussed further on in this session.
If you want to do automatic collation, you have much more of a problem with the tools...

[ ASIDE why auto-collation? A great debate; I find it much faster than manual collation even though it means more transcribed text, and much less error-prone as well. ]

Explosion of transcription tools. No standard yet.
No one has settled, and many of these geared to linking the text to a specific MS image.
Sometimes exactly what you want, sometimes not.
Some of them also serve as publication systems.
Transcription tools

If you want to do automatic collation, you have much more of a problem with the tools...
[ ASIDE why auto-collation? A great debate; I find it much faster than manual collation even though it means more transcribed text, and much less error-prone as well. ]

Explosion of transcription tools. No standard yet.
No one has settled, and many of these geared to linking the text to a specific MS image.
Sometimes exactly what you want, sometimes not.
Some of them also serve as publication systems.
If you want to do automatic collation, you have much more of a problem with the tools...

[ ASIDE why auto-collation? A great debate; I find it much faster than manual collation even though it means more transcribed text, and much less error-prone as well. ]

Explosion of transcription tools. No standard yet.
No one has settled, and many of these geared to linking the text to a specific MS image.
Sometimes exactly what you want, sometimes not.
Some of them also serve as publication systems.
If you want to do automatic collation, you have much more of a problem with the tools...
[ ASIDE why auto-collation? A great debate; I find it much faster than manual collation even though it means more transcribed text, and much less error-prone as well. ]

Explosion of transcription tools. No standard yet.
No one has settled, and many of these geared to linking the text to a specific MS image.
Sometimes exactly what you want, sometimes not.
Some of them also serve as publication systems.
If you want to do automatic collation, you have much more of a problem with the tools...

[ ASIDE why auto-collation? A great debate; I find it much faster than manual collation even though it means more transcribed text, and much less error-prone as well. ]

Explosion of transcription tools. No standard yet.
No one has settled, and many of these geared to linking the text to a specific MS image.
Sometimes exactly what you want, sometimes not.
Some of them also serve as publication systems.
If you want to do automatic collation, you have much more of a problem with the tools...

[ ASIDE why auto-collation? A great debate; I find it much faster than manual collation even though it means more transcribed text, and much less error-prone as well. ]

Explosion of transcription tools. No standard yet. 
No one has settled, and many of these geared to linking the text to a specific MS image. 
Sometimes exactly what you want, sometimes not. 
Some of them also serve as publication systems.
If you want to do automatic collation, you have much more of a problem with the tools...

[ ASIDE why auto-collation? A great debate; I find it much faster than manual collation even though it means more transcribed text, and much less error-prone as well. ]

Explosion of transcription tools. No standard yet.
No one has settled, and many of these geared to linking the text to a specific MS image. Sometimes exactly what you want, sometimes not.
Some of them also serve as publication systems.
And then once you’ve transcribed, how to collate? Here was the gold standard for a decade [TRANS] but it has died a death
And then once you’ve transcribed, how to collate? Here was the gold standard for a decade but it has died a death
There are some current tools, but they each have limitations.
[describe Juxta]
[describe CollateX]

Sad truth: lots of duplication of effort here. I’ve written a collator, so have others; no incentive and no time to make them usable by third parties.
Collation (automatic)

There are some current tools, but they each have limitations.
[describe Juxta]
[describe CollateX]

Sad truth: lots of duplication of effort here. I’ve written a collator, so have others; no incentive and no time to make them usable by third parties.
There are some current tools, but they each have limitations.
[describe Juxta]
[describe CollateX]

Sad truth: lots of duplication of effort here. I’ve written a collator, so have others; no incentive and no time to make them usable by third parties.
Now we get to the fun part (at least I think): you’ve gathered the mss, you’ve generated the data, now you get to analyze it. The boring part is behind you.
Lots of discussion of stemmatic analysis, whether it can be done in an automated way, and what its limitations are.

Very very early state of experimentation with this; no formal theory; people try their methods with their texts and justify their results if the results seem sound.

[EXPLAIN Leuven project]
Stemmatic analysis

In keeping with very–early state of statistical analysis, we have very–hard–to–use tools and processes.
- Get your data file (from CTE, from collator, from friendly biologist helper)
- Stick it in a biological statistical package e.g. Mobyle; get an incomprehensible results file back
- Stick that results file into a visualizer meant for scientists, get your tree
- ...but then it is totally up to you how you interpret that tree. What fun to be an editor.
Stemmative analysis

Input data

In keeping with very–early state of statistical analysis, we have very–hard–to–use tools and processes.
- Get your data file (from CTE, from collator, from friendly biologist helper)
- Stick it in a biological statistical package e.g. Mobyle; get an incomprehensible results file back
- Stick that results file into a visualizer meant for scientists, get your tree
- ...but then it is totally up to you how you interpret that tree. What fun to be an editor.
Stemmatic analysis

Input data
Online statistical tool(s)

In keeping with very-early state of statistical analysis, we have very-hard-to-use tools and processes.
- Get your data file (from CTE, from collator, from friendly biologist helper)
- Stick it in a biological statistical package e.g. Mobyle; get an incomprehensible results file back
- Stick that results file into a visualizer meant for scientists, get your tree
- ...but then it is totally up to you how you interpret that tree. What fun to be an editor.
Stemmatic analysis

In keeping with very–early state of statistical analysis, we have very–hard–to–use tools and processes.
- Get your data file (from CTE, from collator, from friendly biologist helper)
- Stick it in a biological statistical package e.g. Mobyle; get an incomprehensible results file back
- Stick that results file into a visualizer meant for scientists, get your tree
- ...but then it is totally up to you how you interpret that tree. What fun to be an editor.
Editorial analysis

You have your collation, you (hopefully) have your stemma, now it is time to do the fun part – actually editing.

Sad, sad lack of tools for this. Implicit in CTE, which required manual collation; what if you have the automatic collation? Method seems so new that I don’t even know what others do. I wrote my own program...
- displayed each variant with surrounding context
- let me mark spelling & orthographical variants, kept track and kept me consistent
- let me choose a lemma from the available options, or let me emend with an explanation
- from this, my apparatus criticus was automatically created.
You have your established text, you have your mss and notes, you have your stemma; now what? Publication.
- [TODO] mentioned that CSCO doesn’t yet do digital publication, but afaik almost no publisher does, and when they do it is usually a clunky closed system.
- Thus, digital publications to date tend to be self-publications.  
[TRANS] Been going on since at least the 1990s in Armenian studies. A very (technologically) simple example here.
You have your established text, you have your mss and notes, you have your stemma; now what? Publication.
- [TODO] mentioned that CSCO doesn’t yet do digital publication, butafaik almost no publisher does, and when they do it is usually a clunky closed system.
- Thus, digital publications to date tend to be self-publications.
[TRANS] Been going on since at least the 1990s in Armenian studies. A very (technologically) simple example here.
[TRANS] Anastasia / SDPublisher, from B’ham, is another option. Leiden database; still best collection of digital Armenian philological resources. Same system can produce full editions with mss, e.g. Miller’s tale. I’m not aware of other standalone publication options.
[TRANS] Anastasia / SDPublisher, from B’ham, is another option. Leiden database; still best collection of digital Armenian philological resources. Same system can produce full editions with mss, e.g. Miller’s tale. I’m not aware of other standalone publication options.
[TRANS] Anastasia / SDPublisher, from B’ham, is another option. Leiden database; still best collection of digital Armenian philological resources 
Same system can produce full editions with mss, e.g. Miller’s tale 
I’m not aware of other standalone publication options.
Online publication has its own issues of course...[TRANS and discuss]
Publication

• How do we publish?

Online publication has its own issues of course...[TRANS and discuss]
Publication

• How do we publish?
• How do we review?

Online publication has its own issues of course...[TRANS and discuss]
Publication

- How do we publish?
- How do we review?
- How do we ensure their stability / sustainability?

Online publication has its own issues of course...[TRANS and discuss]
Full list of tools mentioned at http://www.interedition.eu/wiki/index.php/Existing_Tools

Please suggest additions!

Thank you / the end. Here’s that wiki page again, to be linked also to COMSt wiki as soon as I can manage; what tools are useful to you?