

is common practice to rearrange the contents of the TEI file using a tool like XSLT as part of the process of converting it into a publishable form, and this rearrangement (along with other aspects of the transformation) does not normally need to be reversible. The published digital surrogate is a lossy derivative of the source document, but this does not present us with any problems as long as display is its only function.

Consider, for example, the following transcription of a Greek ostrakon from the 2nd or 3rd century CE (from <http://papyri.info/ddbdp/o.heid;505>):

DDbDP transcription: o.heid.505 [xml]

[-ca.?-] (*) καὶ ((unintelligible)) (*) κ. ἀπὸ λόγου

[-ca.?-] αὐτῶ... τῶν

[-ca.?-] κβ..... αὐ

[-ca.?-] μου vac.?

5 [-ca.?-] καὶ(*) ὑπ(ἐρ) κ...

[-ca.?-] α(*) Θῶθ

Apparatus

^ 1. or [-ca.?- Νό](του)(?)

^ 1. or Λι(βός)(?) (or (δεκανός)(?))

^ 5. or κ (ἔτους)

^ 6. or λ

In the publication display, line 5 has the word καί (and), but the editor notes that the actual reading might instead be κ (the number 20) and the abbreviation for ἔτους (years). The markup for this line looks like:

```
<lb n="5"/>
```

```
<gap reason="lost" extent="unknown" unit="character"/>
```

```
<app type="alternative">
```

```
<lem>καὶ</lem>
```

```
<rdg>
<num value="20">κ</num>
<expan>
<ex>ἔτους</ex>
</expan>
</rdg>
</app>
<expan>ὕπ<ex>ἔρ</ex></expan>
<unclear>κ</unclear>
<gap reason="illegible" quantity="4" unit="character"/>
```

Our example shows how notes on textual variation tend to be encoded at the locus of variation, but displayed as end- or footnotes. Encoding variations where they occur is easier for the person doing the encoding, because it reduces the possibility of errors arising from having to keep track of changes in two locations while editing, because it removes the need to create links between the locations and because the flow of transcription is not interrupted—the main text and its variations can be recorded in a single pass. The result of this (perfectly natural) bias towards the encoder is that the editor/encoder and the reader/consumer will actually have different mental models of the document. The former will experience variation as a literal fork in the text stream, while the latter will see it as an annotation pointing back at the text.

Readers of printed critical editions are used to seeing critical apparatus as the (highly compressed) set of notes on the text you find at the bottom of the page. These notes are meant to help explain how the editor decided on printing the main text on the page and how the different sources they used to construct that text support or differ from what they printed. The majority of the entries in an apparatus are variant readings from the manuscript tradition and related texts which the editor thinks are significant. Editorial conjectures, or the acknowledgement of a prior editor's conjecture accepted into the text are another. The app. crit. is important, because it shows (at least some of) the evidence for the text as reconstructed by the editor. Despite its scholarly importance, however, the app. crit. is the part that is almost always missing in digital versions of texts.

The reasons for this are various: the text itself may not be copyrightable, but the

app. crit. is usually considered to be the portion of a critical edition that is intellectual property and subject to copyright, so a digital copy of a modern edition may omit the app. crit. for legal reasons. Secondly, due to the extreme compression of the app. crit., it is hard to transcribe accurately and usefully. Properly doing it digitally would entail a certain amount of expansion and interpretation, work that would have to be done by a trained scholar, and would therefore mean a more extensive project than simple digitization. Thirdly, a number of the scholars responsible for producing corpora of digital texts feel that the app. crit. is an artifact of print and unnecessary in the "digital age". We'd be better off, they argue, with digitized versions of every variant manuscript and scholarly edition, which we could align and then compute the textual variances and automatically produce new and better editions or have a multitext in which every variation was available. Finally, I will argue that the TEI's critical apparatus infrastructure (as currently constituted) is insufficient for (and perhaps even inimical to) the encoding of critical apparatus as it commonly appears in print.

Recent developments, however, including improved mechanisms for displaying TEI in the browser and the ongoing revisions to the TEI Pointers specification, point to a possible way towards the re-introduction of the app. crit. to existing digital editions. If we re-conceive the app. crit. as a set of annotations on a text—as it is represented in print—then we can begin to imagine an annotation system which could be used to enhance existing TEI editions by adding to them the information about source variances that they lack. Since the vast majority of these are simply facts (e.g. ms. F has "latebit" where the text prints "tacebo"), they are not subject to any IP restrictions. The wholesale copying of an app. crit. might be a problem, but the facts recorded therein are just facts.

I will conclude with a discussion of the requirements of an annotation system capable of adding critical apparatus notes to existing TEI texts, including discussion of how the TEI Guidelines may need to be adapted to suit the apparatus-as-annotation model. Finally, will demonstrate a prototype system based on published TEI texts from the Perseus project.

Bibliography

Cayless, Hugh A. "Rebooting TEI Pointers," *Journal of the Text Encoding Initiative*[Online], Issue 6 December 2013, Online since 26 September 2013, connection on 30 April 2014. URL: <http://jtei.revues.org/907> ; DOI: 10.4000/jtei.907

Margoni, Thomas and Perry, Mark, "Scientific and Critical Editions of Public Domain Works: An Example of European Copyright Law (Dis)Harmonization," *Canadian Intellectual Property Review*, Vol. 27, p. 157 (November 18, 2011). URL: <http://ssrn.com/abstract=1961535>

Perseus Project Texts, URL: <https://github.com/PerseusDL/canonical>

Walsh, John, Grant Simpson, and Saeed Moaddeli, TEI Boilerplate, URL:
<http://dcl.ils.indiana.edu/teibp/>