Minutes of the standoff-workshop at Humboldt Universität zu Berlin, in January 2014

Present: Andreas Witt, Bertrand Gaiffe, Carolin Odebrecht, Javier Pose, Laurent Romary, Patrice Lopez, Piotr Bański, Peter Stadler (minutes)

The Berlin workshop on standoff annotation was called to take care of the according feature request <http://sourceforge.net/p/tei/feature-requests/378/>. This FR has already been discussed at some Council meeting and it has already been agreed on a separate „standoff“ element (not on the exact name, though) as a sibling to <text>. The task of this workshop was to develop this particular element regarding its definition, its placement, and its content model. Some previous work had already been carried out resulting in some papers [1, 2] which has been summarized by Javier at the beginning of the workshop [3].

Since the participants of the workshop had differing backgrounds, different understandings of *standoff* had to be taken into account: from very small annotations to a text („post-its on a document“) to rather independent documents vaguely related to one or several texts (e.g. personographies or bibliographies). Additionally, these standoff documents could be the result of automated tools (e.g. POS tagger output) or hand-crafted.

The major issue had been that of metadata: Can all metadata concerning the standoff element be captured within the teiHeader or is there the need for a separate metadata section with the *standoff*-element? If all metadata (concerning text and standoff) was merged within the teiHeader, how can the standoff-section be made available within other contexts? If there was a separate metadata section within the standoff element, doesn't that mimic the current functions of teiHeader and text?

Another 'hot' topic had been grouping and nesting of the proposed *standoff*-element. Again, metadata was a big issue in terms of inheritance which led to the general idea of seeing each metadata section of a given standoff-element as a complement to its parent metadata section. Consequently, the teiHeader as the topmost metadata section of a TEI document *can* be complemented by the standoff metadata section but does not need to be, leading to the simplest structure

<TEI xmlns="http://www.tei-c.org/ns/1.0" xmlns:so="http://standoff.proposal">

<teiHeader>

…

</teiHeader>

<so:stdf>

<so:annotations>

…

</so:annotations>

</so:stdf>

</TEI>

where the teiHeader is the metadata section of the standoff element, creating an autonomous annotation document.

In turn, a more complex structure can be created by nesting <so:stdf> elements and adding an optional <so:soHeader>. Concerning the content model of <so:soHeader> it was considered that most of the teiHeader components would be useful, but all of them needed to be optional. Hence, biblFull served as a model.

We started working on an ODD file right during the workshop which is made available at [4], so I refuse to elaborate further on content models etc. but redirect you to sourceforge for further details.

Additional discussions (that should be mentioned) arose around the issue of using TEI elements for creating a standoff structure of a base text, e.g. marking a <lg> just by reference to the appropriate lines rather than injecting that <lg> tag into the text. Ideas had been to create a <mapStruct> element and use a different namespace in analogy to <egXML> but that wasn't brought to an end and the elementSpec in the current ODD file can be considered a stub.

[1] P. Bański (2010). Why TEI standoff annotation doesn’t quite work: and why you might want to use it nevertheless. In Proceedings of Balisage: The Markup Conference, 2010. Vol. 5 of Balisage Series on Markup Technologies

[2] P. Boot (2009). Mesotext: Digitised Emblems, Modelled Annotations and Humanities Scholarship. Pallas Proefschriften

[3] A Generic Formalism for Encoding Standoff annotations in TEI, <http://sourceforge.net/p/tei/feature-requests/_discuss/thread/1e0e4acb/13db/attachment/A%20Generic%20Formalism%20for%20Encoding%20Standoff%20annotations%20in%20TEI.pdf>

[4] Proposal for encoding standoff annotations in TEI, <http://sourceforge.net/p/lingsig/code/HEAD/tree/branches/standoff/proposal.odd.xml>