

Works, Documents, Texts and Related Resources for Everyone

Robinson, Peter

p.m.robinson@bham.ac.uk
Institute for Textual Scholarship, University of Birmingham

Meschini, Federico

fmeschini@dmu.ac.uk
Centre for Textual Studies, De Montfort University

A common trope in discussions of scholarly editions in digital form is to praise, on the one hand, the extraordinary potential of electronic editions while, on the other hand, regretting that so few actual electronic editions come anywhere near realizing this potential (Robinson 2005). The potential is well-known: an explicit hyper-textual structure, publication in a distributed network environment, escape from the storage limit of the printed medium and possession of multiple layout possibilities (such as normalized and diplomatic transcriptions juxtaposed to facsimile images).

The difficulties are also well-known: among them, the need for a formal, comprehensive and efficient encoding scheme to underpin scholarly editions in electronic form. The Text Encoding Initiative Guidelines provided a crucial element, by supplying namings, specifications and structure for key components of electronic editions: thus the specialized lower-level elements for manuscript description and critical apparatus, along with higher-level elements such as `msDescription` and `facsimile`. However, the TEI does not address two areas, crucial for the full encoding of scholarly editions in electronic form:

1. The naming of components of the editions: thus, of the works edited and their parts; the source manuscripts or print documents and their parts which carry the texts of the work edited;
2. The relationships between the components: thus, between the documents, the texts they

carry, and the works which those texts instance.

This paper reports on a scheme prepared by the authors, designed to provide a solution to the problems proposed in both areas. The provision of a shared epistemological framework for handling works, texts and text sources (cf. Buzetti 2009) will also facilitate the shift from stand alone publishing frameworks to shared distributed on-line environments, enabled by powerful and flexible underlying infrastructures,¹ generally named Virtual Research Environments (Fraser 2005, Dunn et al. 2008).

This framework will advance interoperability, long a problem area in electronic texts. Interoperability has been defined by IEEE as “The ability of two or more systems or components to exchange information and to use the information that has been exchanged”.² A recent briefing paper by Gradmann identifies four different levels of interoperability, one built on the top of the other. From the bottom these levels are technical/basic, syntactic, functional and semantic. While technologies such as TCP/IP, HTTP and XML already provide sound basis for interoperability at the lower levels, much work is still to be done at the top levels. The semantic frame for interoperability offered by this scheme speaks to this need.

Semantic issues in networked publication systems are advanced by the work done in the last years on the ‘Semantic Web’ (Berners-Lee et al. 2001), which has recently evolved into the Linked Data initiative (Berners-Lee 2006). The Semantic Web seems to have survived its own hype, having finally entered the plateau of productivity phase, as happened for XML some years ago. The ontological level of the Semantic Web stack, represented by the OWL language, has presented a steep learning curve, due partly to its roots in Description Logic and First-Order Logic (Gruber 1993), but also presents at the same time the greatest potential.

The relationship between textual scholarship in its electronic dimension and ontologies has not hitherto been much apparent, as textual scholars using digital methods have focussed rather on the related, but separated field of Library and Information Science (Vickery 1997). However, ontologies have much

to offer the textual editing enterprise. Both ‘recensio’ and the construction of a stemmatic graph are implicit formalizations that would benefit from the adoption of an explicit modelling. Moreover, both Sperberg-McQueen and Peter Shillingsburg implicitly hints at the potentialities of an ontological approach in scholarly editions, the former when writing about the “infinite set of facts related to the work being edited” (Sperberg-McQueen 2002) and the latter about “electronic knowledge sites” (Shillingsburg 2006).

In the world of digital humanities and electronic editions proficient uses of ontologies have already appeared, such as the Discovery³ and the Nines⁴ projects, also leveraging existing standards from related sectors such as IFLA’s FRBR⁵ (Mimno 2005) or the cultural heritage oriented CIDOC-CRM⁶ (Ore et al. 2009).

Substantial work is now being done on implementing an actual interchange and interoperability framework for electronic editions, and arbitrary portions of them, of the kind, in (for example) the COST Action Interedition.⁷ A first proposal by Peter Robinson (Robinson 2009) was based on the Kahn/Wilensky Architecture (Kahn et al. 1995),⁸ having therefore a naming authority together with a series of key/value pairs identifying portions of an electronic text, which therefore could be exchanged over the net thanks to a protocol such as the one established by the OAI-PMH standard.⁹ This addressed the first need stated above, for agreed conventions on naming. The second need, for formal expression of relationships, is addressed by the adoption of the Linked Data paradigm. While keeping the use of the Kahn/Wilensky Architecture for the labelling system, and using a URN-like syntax compatible with the Semantic Web requirements, an ontology representing the entities involved together with their relationships has been developed.

The main entities of this ontology are:

- ‘Work’: Canterbury Tales, and ‘WorkPart’, the first line of the Canterbury Tales;
- ‘Document’, the Hengwrt or the Ellesmere manuscripts, and ‘DocumentPart’, a page, folio or quire, which might carry an instance of the ‘Work’

- ‘Text’: a single instance of a work, or work part, in a document or document part. Thus: the *text* of the *work* ‘The Canterbury Tales’ as it appears in the *document*, the Hengwrt manuscript;

The three-fold distinction between ‘Work’, ‘Document’ and ‘Text’ reflects the fundamental scholarly distinction between the ‘Work’, independent of its realization in any object; the ‘Document’ which might carry an instance of the ‘Work’; and the ‘Text’: the instance of the work in the document. Digital resources such as ‘Image’ or ‘Transcript’ are related to ‘Text’ and ‘Document’ and their parts, using relationship such a ‘hasImage’, ‘isTranscriptOf’, or ‘transcribedFrom’. Basic properties such as “isPartOf” or other properties from existing vocabularies, such as Dublin Core,¹⁰ have also been used, so to guarantee compatibility with other schemes in the best possible way. The resulting RDF can be stored in a triplestore and made available on the web, so to allow further uses from third parties without the need to establish exclusive protocol verbs.

This paper will present the methodological thinking behind the development of this ontology for the interchange of electronic editions of literary texts, starting from the first proposal until the more recent developments. The ontology will be contextualized with the existing related standards, particularly FRBR, CIDOC-CRM and the recent OAI-ORE¹¹ (a gross-grained vocabulary for the reuse and exchange of digital objects developed by the Open Access Initiative) and with the similar initiative of the Canonical Text Service Protocol (CTS),¹² which recently also added an ontological dimension to its basic syntax (Romanello et al. 2009).

References

- Berners-Lee, T., Hendler, J., Lassila, O.** (2001). ‘The Semantic Web’. *The Scientific American*. **May 2001**: 34–43.
- Berners-Lee, T.** (July 2006). *Linked Data*. <http://www.w3.org/DesignIssues/LinkedData.html>.
- Buzzetti, D.** (2009). ‘Digital Editions and Text Processing’. *Text Editing, Print and the Digital*

World. Deegan, M., Sutherland, K. (eds.). Ashgate, pp. 45-61.

Dunn, S., Blanke, T. (2008). 'Next Steps for E-Science, the Textual Humanities and Vres'. *D-Lib Magazine*. **1/2**. <http://www.dlib.org/dlib/january08/dunn/01dunn.html>.

Fraser, M. (2005). 'Virtual Research Environments: Overview and Activity'. *Ariadne*. **44**. <http://www.ariadne.ac.uk/issue44/fraser/>.

Gradmann, S. *INTEROPERABILITY. A key concept for large scale, persistent digital libraries*. www.digitalpreservationeurope.eu/publications/briefs/interoperability.pdf.

Gruber, T.R. (1993). 'A translation approach to portable ontology specifications'. *Knowledge Acquisition*. **5**: 199–220.

Kahn, R., Wilensky R. (May 1995). *A Framework for Distributed Digital Object Services*. <http://WWW.CNRI.Reston.VA.US/home/cstr/arch/k-w.html>.

Mimno, D., Crane G., Jones, A. (2005). 'Hierarchical Catalog Records: Implementing a Frbr Catalog'. *D-Lib Magazine*. **10**. <http://www.dlib.org/dlib/october05/crane/10crane.html>.

Ore, C., Eide, Ø. (2009). 'TEI and cultural heritage ontologies: Exchange of information?'. *Literary and Linguistic Computing*. **2**: 161-172. <http://llc.oxfordjournals.org/cgi/content/abstract/24/2/161>.

Robinson, P. (2005). 'Current Issues in Making Digital Editions of Medieval Texts— or, Do Electronic Scholarly Editions Have a Future?'. *Digital Medievalist*. <http://www.digitalmedievalist.org/journal/1.1/robinson/>.

Robinson, P. (2009). 'Electronic Editions for Everyone'. *Text and Genre in Reconstruction*. McCarty, W. (ed.). Cambridge: Open Book Publishing, pp. 183-201.

Romanello, M., Berti, M., Boschetti, F., Babeu A., Crane G. (2009). *Rethinking Critical Editions of Fragmentary Texts by Ontologies*. Milano, Italy: Elpub. <http://conferences.aepic.it/index.php/elpub/elpub2009/paper/view/158>.

Sperberg-McQueen, C. M. (2002). *How to Teach Your Edition How to*

Swim. <http://www.w3.org/People/cmsmcq/2002/cep97/swimming.xml>.

Shillingsburg, P. (2006). *From Gutenberg to Google: Electronic Representations of Literary Texts*. Cambridge: Cambridge University Press.

Vickery, B. C. (1997). 'Ontologies'. *Journal of Information Science*. **4**: 277-286. <http://jis.sagepub.com/cgi/content/abstract/23/4/277>.

Notes

1. Such as the European initiative DARIAH <<http://www.dariah.eu/>>
2. <<http://en.wikipedia.org/wiki/Interoperability>>
3. <<http://www.discovery-project.eu/>>
4. <<http://www.nines.org/>>
5. <<http://www.ifla.org/en/publications/function-al-requirements-for-bibliographic-records>>
6. <<http://cidoc.ics.forth.gr/>>
7. <<http://www.interedition.eu/>>
8. Which constitutes also the basis for the Handle system <<http://www.handle.net/>>
9. <<http://www.openarchives.org/OAI/openarchivesprotocol.html>>
10. <<http://dublincore.org/>>
11. <<http://www.openarchives.org/ore/>>
12. <<http://chs75.chs.harvard.edu/projects/diginc/techpub/cts>>