

An Ontological View of Canonical Citations

Matteo Romanello, <matteo.romanello@kcl.ac.uk>

Michele Pasin, <michele.pasin@kcl.ac.uk>

Canonical citations are references to Classical (i.e. Greek and Latin) texts that are expressed by scholars by means of an abridged canonical format (1,2). They fulfil the function of providing an abstract reference scheme for texts (somehow similar to the function of geographical coordinates to express references to places) since they allow us to express references to them no matter what particular text edition we are actually looking at. For example, the reference to the twelfth book of Homer's *Iliad* expressed as “Hom. *Il.* XII 1” can be resolved to the text of the same passage as established in various critical editions of that work, where “XII” and “1” are the “coordinates” that allow us to locate that precise text passage within all critical editions of Homer's *Iliad*.

Historically, canonical references are the result of an effort – whose origins can be traced back to the Renaissance (3,4) – made by the scholarly community as a whole to provide a precise, stable and shared way to refer to Classical texts. Since the early stages of Humanities Computing and Digital Humanities (5-7), canonical references were regarded as the ideal candidate on which to experiment the potentialities of hypertext: indeed they can be seen as hyperlinks *in potentia* pointing a text from within another. More recently (8) they were considered as discipline-specific kind of named entities that Classics scholars should be provided with tools to search for within their texts.

J.D. Bolter describes classical philology as “the art of explicating an ancient text by exploring its relationships to other specific texts and to the corpus of ancient literature as a whole”. In such a discipline the act of *referring* to texts – that J. Unsworth has listed among the “scholarly primitives” (9) – becomes even more crucial than in other disciplines since texts are the very research objects of classical philology and references to them play a key role in constructing argumentations.

The point we want to make in this paper – that follows up on one made already by N. Smith (10) – is that since citations reflect already an ontological view of texts in this specific domain, and specifically how classicists perceive ancient texts as objects, formalising them as an ontology will allow us to more fully capture their meaning and exploit the information they carry. We claim that the specification of such an ontology is worthwhile for at least the following reasons:

- 1) it allows us to disentangle from an ontological point of view the complex relationships between, for instance, a canonical reference found in a journal paper and the manuscripts and editions of the text that we can access via that reference;
- 2) it allows us to define types of references and alternative representations of the same reference: this is an important step towards tools that allow automatic formatting of such references according to various styles (as it happens already for modern bibliographic references with Zotero);
- 3) it provides us with a way to access the meaning of canonical references beyond their surface appearance, which might vary substantially as in the case of “Hom. *Od.* I 1” and “α 1”, two canonical references to the same passage but conforming to different citation styles.

However, none of the existing ontologies actually model the deep meaning of canonical references. An interesting attempt to formalise citation by means of an ontology is CITO (11) which however looks exclusively at modern bibliographic references and focuses in particular on citation types. As it was observed already (10,12), the distinction made by FRBR (Functional Requirements for Bibliographic Records) between a work, its expressions and its manifestations can be adapted to represent texts in the Classics domain as well. In this paper we propose an initial implementation of an canonical reference ontology based on FRBRoo which is the result of a process aimed at harmonising FRBR with the CIDOC-CRM (13).

A key aspect we have to face is to determine at which ontological level of the cited object a canonical

reference is pointing to. A citation such as “Hom. *Od.* I 1” is it referring to the abstract notion of *Odyssey* (i.e. a *work* in the FRBR model) or to a particular version (e.g. edition, translation, etc.) of that work (i.e. a FRBR expression)? It might help to observe that this reference can be solved by a human reader for example into the text of that passage in French translation: therefore it is not being specified at the expression level. The textual coordinates of the citation, namely “first line of the first book”, expressed by the string “I 1” clearly refer to a logical citation scheme that applies already to the abstract notion of *Odyssey* (i.e. a FRBR work). Thus we can say that a canonical citation follows a given citation scheme that characterises a particular literary text and might differ from one to another. That citation scheme is a conceptual object and is the result of the work of scholars to guarantee the ability of citing literary texts.

To illustrate the notion of *logical* citation scheme as opposed to a *physical* one let us examine a single case, that is the Athenaeus' *Deipnosophistae*. Scholars cite this work by means of canonical references that follow a logical citation scheme derived from a physical one (e.g. “Ath. *Deipn.* XV 694 e-f”). The textual coordinates “694 e-f” refer to the pagination of the *editio princeps* (i.e. first edition) of the text by Isaac Casaubon dated 1598, and specifically to sections “e” to “f” of page 694. At first it seems a physical citation scheme. But since all editions after Casaubon's provide the readers with marginal numbers referring to that pagination it became a logical citation scheme: indeed 694 does not refer anymore to a physical page within more recent editions such as Olson's.

Canonical references are defined by the practice of scholars or, borrowing the words from the FRBRoo specification, they are “cultural constructs” and “as such, they have a context, a history, and a use in time and space by some groups of users”. Given all these reasons, we propose to introduce a class Citation by subclassing the CIDOC-CRM's class “E41 Appellation” and a class Citation_Scheme subclassing the class “E28 Conceptual Object”. Since we argued already how canonical citations and geographical coordinates play a similar function, it is worth to note that given this definition they both happen to share the same superclass. Further work is then required in order to extend this conceptual model so that it can support more complex reasoning tasks, such as translation mechanisms among different citation schemes, or automated extraction of citations from non structured materials. In fact, as it happens with spatial coordinates schemes, CIDOC-CRM provides only a generic framework usable for instantiating such entities; in order to have domain-specific mechanisms such as the ones above suitable extensions need to be manually developed.

To sum up, in this paper we describe the implementation of an ontology to model canonical references that builds upon the solid conceptual models already defined by CIDOC-CRM and FRBRoo. In the framework of a Classics cyberinfrastructure (8), such an ontology is meant to support the interoperability of tools that are being currently developed to extract (14), retrieve (10) and solve (15) canonical references.

References:

1. Romanello M. A semantic linking framework to provide critical value-added services for E-journals on classics [Internet]. In: Mornati S, Chan L, editors. ELPUB2008. Open Scholarship: Authority, Community, and Sustainability in the Age of Web 2.0 - Proceedings of the 12th International Conference on Electronic Publishing held in Toronto, Canada 25-27 June 2008. 2008 [cited 2008 Aug 11]. p. 401-414. Available from: http://elpub.scix.net/cgi-bin/works/Show?401_elpub2008
2. Romanello M. A Semantic Linking System for Canonical References to Electronic Corpora [Internet]. In: Zemanek P, editor. International Conference on Electronic Corpora of Ancient Languages : proceedings of the international conference, Prague, November 16-17, 2007. Prague: 2007. p. 107–120. Available from: <http://eprints.rclis.org/16239/1/Romanello2008.pdf>
3. Martin H. Du livre à la lecture [Internet]. In: Jacob C, editor. Des Alexandries II. Les métamorphoses du lecteur. Bibliothèque nationale de France; 2003 [cited 2010 Oct 26]. p. 35-

45. Available from: <http://hal.archives-ouvertes.fr/hal-00131623/en/>

4. Berra A. Manier le thésaurus grec. In: Jacob C, editor. Les main de l'intellect. Paris: Albin Michel; 2011 (forthcoming).
5. Bolter JD. Hypertext and the Classical Commentary. In: Accessing antiquity : the computerization of classical studies. Tucson: University of Arizona Press; 1993. p. 157-171.
6. Crane G. From the old to the new: intergrating hypertext into traditional scholarship [Internet]. In: Proceedings of the ACM conference on Hypertext. Chapel Hill, North Carolina, United States: ACM; 1987 [cited 2009 Feb 2]. p. 51-55. Available from: <http://doi.acm.org/10.1145/317426.317432>
7. McCarty W. Humanities Computing. Palgrave Macmillan; 2005.
8. Crane G, Seales B, Terras M. Cyberinfrastructure for Classical Philology. Digital Humanities Quarterly [Internet]. 2009 [cited 2010 Jul 19];3(1). Available from: <http://www.digitalhumanities.org/dhq/vol/3/1/000023/000023.html>
9. Unsworth J. Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this? [Internet]. 2000; Available from: <http://www3.isrl.illinois.edu/~unsworth/Kings.5-00/primitives.html>
10. Smith N. Citation in Classical Studies. Digital Humanities Quarterly [Internet]. 2009 Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure Winter [cited 2009 Mar 15];3(1). Available from: <http://www.digitalhumanities.org/dhq/vol/003/1/000028.html>
11. Shotton D. CiTO, the Citation Typing Ontology. Journal of Biomedical Semantics [Internet]. 2010 [cited 2010 Oct 25];1(Suppl 1):S6. Available from: <http://dx.doi.org/10.1186/2041-1480-1-S1-S6>
12. Mimno D, Crane G, Jones A. Hierarchical Catalog Records. D-Lib Magazine [Internet]. 2005 10 [cited 2010 Jul 22];11(10). Available from: <http://www.dlib.org/dlib/october05/crane/10crane.html>
13. Doerr M, LeBoeuf P. Modelling Intellectual Processes: The FRBR - CRM Harmonization [Internet]. In: Thanos C, Borri F, Candela L, editors. Digital Libraries: Research and Development. Springer Berlin / Heidelberg; 2007. p. 114-123. Available from: http://dx.doi.org/10.1007/978-3-540-77088-6_11
14. Romanello M, Boschetti F, Crane G. Citations in the Digital Library of Classics: Extracting Canonical References by Using Conditional Random Fields [Internet]. In: Proceedings of the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries. Suntec City, Singapore: Association for Computational Linguistics; 2009. p. 80–87. Available from: http://portal.acm.org/ft_gateway.cfm?id=1699763
15. Ruddy D, Rebillard E. Text Linking in the Humanities: Citing Canonical Works Using OpenURL [Internet]. 2009 [cited 2009 Sep 11]; Available from: <http://www.cni.org/tfms/2009a.spring/abstracts/PB-text-ruddy.html>