Citations and Annotations in Classics: Old Problems and New Perspectives

Matteo Romanello
King's College London
Department of Digital Humanitiesd
26-29 Drury Lane
London WC2B 5RL
matteo.romanello@kcl.ac.uk

Michele Pasin
Nature Publishing Group
The Macmillan Building
4 Crinan Street
London, N1 9XW, UK
michele.pasin@nature.com

ABSTRACT

Annotations played a major role in Classics since the very beginning of the discipline. Some of the first attested examples of philological work, the so-called *scholia*, were in fact marginalia, namely comments written at the margins of a text. Over the centuries this kind of scholarship evolved until it became a genre on its own, the classical commentary, thus moving away from the text with the result that philologists had to devise a solution to *linking* together the *commented* and the *commenting* text. The solution to this problem is the system of canonical citations, a special kind of bibliographic references that are at the same time very precise and interoperable.

In this paper we present HuCit, an ontology which models in depth the semantics of canonical citations. Its main goal is to underpin a Knowledge Base providing the information that an automatic expert system requires in order to read and interpret correctly this type of citations. Finally, we describe how HuCit can be combined together with the OAC ontology to publish in a semantic format a specific kind of machine-generated annotations, that is annotations concerning the canonical citations contained in secondary sources (e.g. journal articles, commentaries, etc.).

Categories and Subject Descriptors

[**Computing methodologies**]: Ontology engineering

Keywords

ontology, canonical citations, digital classics, knowledge base

1. INTRODUCTION

Classics as a discipline has always been heavily concerned with annotations since its very beginnings. Some of the first attested examples of philological work, the so-called *scholia*, were in fact marginalia, namely comments written at the margins of a text [8]. Over the centuries this kind of scholarship evolved until it became a genre on its own, the *classical commentary*, thus moving away from the text with the result that philologists had to devise a solution to linking together the *commented* and the *commenting* text [1, 4].

To solve this problem scholars in Classics have devised–well before the advent of digital technologies–a specific kind of references that are called *canonical citations*. Such a solution had to be at the same time precise and interoperable: *precise* because texts are the fundamental object of philological research, therefore a scholarly discourse about texts needs a very accurate and unambiguous way of referring to them; *interoperable* because although texts may exist in different editions and translations, scholars need to be able to refer to specific sections of them without having to worry about the many possible variations in pagination or layout each single edition may present. Some examples of caonical citations are "Aristot. *Poet.* 1451a.35", "Hom. *Il.* I 1-10" and "Thuc. I 33.1" that refer to specific sections respectively of Aristotle's *Poetics*, Homer's *Iliad* and Thucydides' *Histories*.

Although some ontologies that include the notion of citation already exist, none of them is really suitable to model the specificities of citations in the Humanities–and specifically in Classics. To fill this gap we developed HuCit, an ontology that aims at representing in greater detail the semantics of canonical citations. The starting point for our work was trying to answer questions such as "what is the denotation of a canonical citation?", "how is a canonical citation different from a normal citation?". These are important issues to solve when trying to construct a computational framework that allows to extract, store and interrogate the network of canonical citations that are commonly used in the classics literature.

HuCit was developed with a two-fold goal in mind: on the one hand, to support the automatic extraction of canonical citations from texts and specifically by providing a source of domain knowledge that can be used to validate and/or disambiguate canonical citations; on the other hand, to publish the so extracted citations in a machine-readable format that is also compliant with the Semantic Web and Linked Open Data (LOD) philosophy [5].

This paper is structured as follows: in section 2 we give an

overview of related work. In section 3 some examples of canonical citations are described in greater detail in order to highlight what are their specificities and to explain how they function. The ontology itself is introduced in section 4 where the main choices in terms of design are discussed in light of its intended use. In section 5 we provide a practical example of how a Knowledge Base built upon HuCit can be used to support the automatic extraction of canonical citations. Finally we illustrate how HuCit in combination with the Open Annotation Collaboration (OAC) ontology can be used to store and publish as LOD the citation network which starts to emerge once citations between primary and secondary sources are captured.

2. RELATED WORK

2.1 Citations in Formal Ontology

In general, we can identify two main approaches to modeling bibliographic citations. The first one stresses the performative aspect of a citation, that is, it considers it as the *act* of linking two documents by means of a scholarly relationship. The second one instead focuses more on the *textual* dimension of a citation, namely, it considers it as a series of symbols that can be interpreted as a pointer to another bibliographical entity. It is important to underline that the two approaches are not antithetical, in fact both often coexist within a single data model or ontology. However this duality can generate some terminological ambiguity since people may refer to one or the other aspect using the same terms, which most often are '*citation*' and '*reference*'. We will look at both approaches in turns.

Citations as performative entities can be encoded simply as a relation between two bibliographical objects, or indirectly via a reified entity that represents the 'reference event' - the act of referencing another object. We found no evidence of the latter strategy, on the contrary there are many instances of the former one. For example the BIBO ontology uses the *cites* object property that relates "a document to another document that is cited by the first document as reference, comment, review, quotation or for another purpose"[1]. In AKT the *cites-publication-reference* is used very similarly[2]. More interestingly, ontologies like SWAN [2] and CiTO [12][3] go one step further in this direction by providing an explicit categorization of the types of rhetotical actions the citing relationship could mean. So for example in CiTO one can specify whether a citation '*cites-as-evidence*' or '*critiques*' another object, thus allowing the creation of networks of scholarly bibliographical relationships that are semantically very rich. The main limitation of this approach, however, is that since the ontological commitment of these citation relations rely almost entirely on their scope notes (no formal semantics is provided) it is very hard to achieve consensus with regards to the meaning of these relationships (as different people may mean different things by using the same relationship term).

Let us now consider the second approach to modeling citations. In this case the focus is on the specific form a citation takes in the main body of a paper or within the references section (for this reason, it is often called *reference*). In other words, the accent here is posed on symbolic level of a citation: what text it contains, how it is structured or how it is ordered. A citation object, thus intended, plays the same role of an address: it gives you some information useful for finding an article. So for example in BiRO [4] a 'Bibliographic Reference' is seen as a textual component, which is normally part of a 'Reference List'. A similar approach is taken in DoCO[5] and DEO[6] where a *doco:BibliographicReferenceList* is said to contain one or more *deo:BibliographicReference*. Arguably the most comprehensive formalization of this idea is the one provided by Gruber in his Bibliographic-Data ontology [7], which predates the era of Linked Data (in fact it was not implemented using any of the RDF family of languages) but contains a number of useful insights. In particular, Gruber pays attention to the issue of clarifying the terminological ambiguity between *citation* and *reference* by observing that a "reference is distinguished from a citation, which occurs in the body of a document and points to a reference" and that a "bibliographic reference is a description of some publication that uniquely identifies it, providing the information needed to retrieve the associated document". Based on these central ideas he provides a detailed characterization of subclasses of *Publication-Reference*, such as *Article-Reference*, *Book-Reference*, *Book-Section-Reference* and so on, to the aim of enabling "the development of reference-formatting styles that are independent of database or tool, and to support remote services such as bibliography database search and reference-list generation". In our own work we share much of Gruber's approach, especially the conceptual and terminological distinction between a reference - the text appearing in the references section of a document - and a citation - the text appearing in the body of the document. Canonical citations, as we will see, are a special case of citations for they also normally appear within the main body of a text.

2.2 Citations and Annotations in Digital Classics

In carrying out this work we did not start from scratch but we built upon the results of research that was carried out over the last decade within the Digital Classics community with the goal of replicating the system of canonical citations in a digital, networked environment.

A first strand of research in this area, characterised by a stronger focus on how to transform canonical citations into a machine-actionable links on the Web, has produced two different solutions: CWKB OpenURL format [8] and the Canonical Text Services (CTS) protocol[9] [14]. The two solutions are very similar to each other as the problem they try to tackle is essentially the same, but at the same time they shed light on it from two different perspectives: a more librarian

---

[1] http://bibliontology.com/

[2] http://www.aktors.org/publications/ontology/

[3] http://www.essepuntato.it/lode/http:/purl.org/spar/cito

[4] http://vocab.ox.ac.uk/biro#term_references

[5] http://www.essepuntato.it/lode/http://purl.org/spar/doco#d4e224

[6] http://www.essepuntato.it/lode/http://purl.org/spar/deo#d4e131

[7]

[8] http://cwkb.org/matrix/20100922/

[9] http://www.homermultitext.org/hmt-doc/

perspective in the case of the former and a perspective that is closer to the field of digitial scholarly editing in the case of the latter.

Another area of research is concerned with the related problem of how to capture automatically such citations from unstructured texts. Gregory Crane argued that services which perform this task with a high degree of accuracy are an essential part of a cyberinfratructure for research in Classics. From a technical point of view, he suggested that the extraction of citations from texts can be seen as a discipline-specific task of Named Entity Recognition and, more generally, Information Extraction. Along the same line of thought, some recent studies have laid the foundations of a such a service by showing how a machine learning based-system that can be trained to extract citations from multilingual corpora [11, 9].

Finally, it is worth mentioning the Pelagios project[10] as an example of how OAC has been used in the field of Digital Classics. The goal of Pelagios was to create a network of linked resources related to ancient geographical places in order to new modes of discovery and visualisation of geodata [13]. Its approach to the problem is simple yet very powerful and can be summarised in the following basic tenets: 1) resources that make some assertion about the same place of the ancient world are linked to each other and such assertions are treated as OAC annotations; 2) the vocabulary to describe places is drawn from Pleiades, a community-built gazetteer of ancient places[11], and uses URIs to identify places.

3. DOMAIN ANALYSIS

First off, we should say that the use of canonical citations in order to refer to primary sources is part of the training of any classicist. Therefore, a knowledge base built to support the automatic extraction of such citations from texts has to contain a number of notions that are normally acquired by scholars during their training as classicists.

A distinguishing feature of canonical citations is that they are valid no matter what specific edition one uses to look them up. For example, the citation "Hom. *Il.* I 1-10", which identifies the first ten lines of the first book of Homer's *Iliad*, can be looked up in the Italian translation by Monti as well as in the critical edition by M.L. West published in two volumes between 1998 and 2000. However, not all citations in classics are canonical. The way of citing fragmentary texts [10], for example, is not canonical since any citation refers to a specific (critical) edition of the text and there is no way to cite a fragmentary texts without referring to a specific edition.

In general, canonical citations are possible because they refer to a standardized version of a classical text, which is then referenced to by any other subsequent edition. What is *canonical* about such kind of citations is the abstract text structure they refer to, which is often reducible to the physical characteristics of a specific edition (eg division into books, chapters, pages etc.). In this sense *canonical* means agreed

[10]http://pelagios-project.blogspot.com/
[11]http://pleiades.stoa.org/

upon by classicists over centuries and established through the practice.

If we look at the historical evolution of this system, we can highlight two possible cases that led to the selection of a canonical text structure: 1) the text structure is the one originally given to a work by its author: this is the case of most of the poetic compositions with the exception of those originating from an oral tradition, as it might have happened with the Homeric poems; 2) the standard, canonical structure was imposed onto the text later on as an attempt to standardize and make more precise–that is fine-grained–the way of referring to the text. For example when referring to Aristotle's works we usually refer to a structure, the so-called Bekker numbers, that became canonical after Bekker's XIX century edition of the *Corpus Aristotelicum*.

What is meant here by *text structure* is a hierarchical structure of elements into which a given text is divided and organized. The division of each book of the Bible into chapters and verses, for example, happened in different stages: the systematic division of the text into chapters of approximately the same length is due to Archbishop Stephen Langton in the XIII century, whereas the 16th century French printer Robert Estienne is responsible for division into verses that is still adopted in modern editions of the Bible.

It is worth pointing out though that the picture just delineated is not always so clearcut, and there are a number of hybrid situations. An interesting one is the canonical way of citing the works by Aristotle, where canonical and non-canonical text structures can often co-exist. In addition to the standard way of citing his works, based on the page numbers of the 1831 edition by Immanuel Bekker, there is another way which refers to a division of the text into books, chapters and sentences (cfr. fig. 1).

Furthermore, it is common for modern editions of classical texts to refer to a number of overlapping structures of the text. The most simple example is a critical edition of Homer's *Iliad* where one will find at least two co-existing hierarchical structures of the text that can be used to cite it: one is the structure of the edition itself as defined by its pagination and the other is the canonical structure of the text, namely its division into books and poetic lines.

The way in which such overlapping text structures are expressed depends on the medium: in the case of printed editions the references to such structures are part of the paratext (see image above), whereas in the case of a TEI-encoded electronic edition they will be expressed by using markup elements such as `<milestone>` and `<div>`.

Finally, another important characteristic of canonical citations is that, despite the fact that their use has been standardized to a great degree, there exist multiple, equivalent ways of expressing the very same citation. The reason for choosing one or the other citation style may differ: in some cases it is merely matter of style while in other cases may depend on the publication venue or even audience for which someone is writing. For example, in a journal or edited volume with a strong focus on epic poems it is common to find canonical citations to Homer's *Iliad* and *Odyssey* expressed
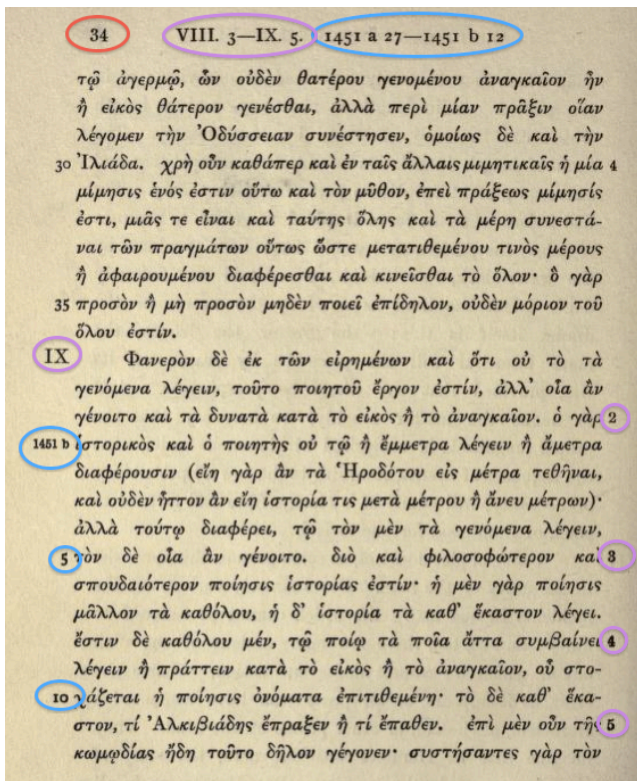
Figure 1: The figure shows page 34 of Butcher's translation (1895) with critical text of Aristotle's *Poetics*. The rich paratext contains references to two different systems to divide the text: the canonical Bekker numbers (highlighted in blue) and the division of the text in chapters and sentences (in purple).

in a particularly concise format (see 3 for an example) which uses the uppercase letters of the Greek alphabet to identify the books of the Iliad and the lowercase ones for the Odyssey. The reason to this is that in such publications the number of references to the homeric poems tends to be in the order of hundreds.

4. ONTOLOGY WALKTHROUGH

HuCit is a light-weight ontology which consists in its current version (0.9.1) of 10 classes and 16 properties[12]. Such a small number of classes is due to the fact that the ontology imports and extends–mostly by defining more specific sub-classes–classes and concepts that are already defined in existing ontologies such as CIDOC-CRM[13] and FRBRoo[14]. In this section we shall introduce the main classes of HuCit and discuss the most important modeling decisions.

4.1 Citation and Canonical Citation

As noted already above, we deliberately treated citations as textual objects rather than performative ones. As a result

---

[12]http://www.essepuntato.it/lode/owlapi/http: //purl.org/net/hucit
[13]http://www.cidoc-crm.org/
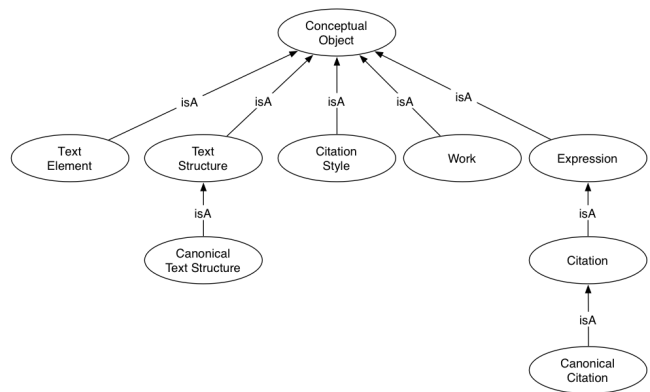[14]http://www.cidoc-crm.org/frbr_inro.html

Figure 2: Overview of classes contained in HuCit.

we have the classes `Citation`, and its sub-class `Canonical-Citation`, which are used to represent occurrences of a citation in the main body of a text. The main reason for this is that we are interested not only in capturing the fact that a citation between two texts exists, but also in capturing the specific structure of the string used to express that citation. Such interest comes mostly from the purpose for which HuCit was designed, that is to build a Knowledge Base to support the extraction of such citations.

Defining from an ontological point of view what a citation is, proved to be more complicated a task than it may seem. As the fact that all classes in HuCit are sub-classes of `ConceptualObject` shows, citations are abstract rather than physical objects. Specifically, we are dealing with what are usually called *representations* or *content-bearing objects*: canonical citations are in fact sequences of characters that are understood by human readers as identifying a specific section of a text. Therefore we followed the approach delineated by Mizogouchi in his ontology of representations [6] and partially investigated by one of the two authors in previous work [7], which led us to ask ourselves what are respectively the *form* and the *content* of a canonical citation.

4.2 Form and Content of a Canonical Citation

Equivalent are two or more citations that express the same content but present different forms, such as for example 'Hom. *Od.* I 1', 'Hom. *Od.* 1.1' and ' 1'. Although what all these citations mean is "book 1, line 1 of Homer's *Iliad*", independently from the reference edition, they differ as to how such meaning is expressed or, in other words, they follow different citation styles.

The styles of canonical citations, however, are less clearly defined and standardized than those for citing modern publications, such as for example the Harvard or MLA style. What should also be noted is that it goes beyond the current scope of our work to define citation styles to such a level of details that one can built on top of our ontology an application to apply automatically different styles to the same canonical citations, as reference manager applications such as Zotero or Mendeley can easily do for modern publications by using sets of rules expressed in the Citation Style Language (CSL) format[15].

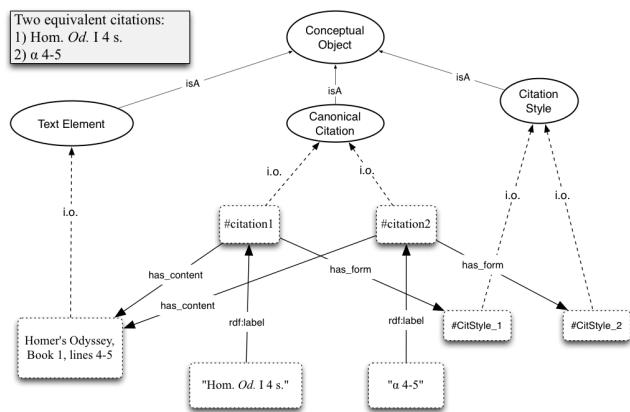---

[15]http://citationstyles.org/

Figure 3: Example of how two equivalent canonical citations can be modeled in HuCit.

4.3 TextStructure and TextElement

More complex yet interesting is to define the *content* of a citation. As was already observed above, two equivalent canonical citations present different citation styles but refer to the same section of a text. What is canonical–that is standard and agreed upon–in such citations is the text structure that is used to refer to a text. Such a structure exists at a more abstract level than the single existing editions of that text. To formulate this by using a terminology drawn from FRBR, canonical citations imply a text structure that is common to all the existing `Expressions` of a given `Work`. Already mentioned examples of canonical text structures are the division of the Bible into chapters and verses or the so-called Bekker numbers for the texts of Aristotle.

A `TextStructure` is a set of hierarchically-ordered `TextElements`: for example, the canonical structure of Homer's *Iliad* is made of books and each book is made of poetic lines. Therefore a part-of relationship is defined between a `TextStructure` and its `TextElements` but also between two or more `TextElements`. The content of a canonical citation is ultimately a list of one or more `TextElement` within a `CanonicalTextStructure`.

The properties `follows` and `precedes` are necessary to retain and make explicit information about the topology of texts, such as the fact that "book 1, line 611" of Homer's *Iliad* follows "book 1, line 610" and precedes "book 2, line 1". A citation to more than one verse is usually expressed as a range, such as in "Hom. *Il.* 1.1-21": in order to transform such a reference into quantifiable information we need to know which lines are comprised between "1.1" and "1.21".

5. HUCIT IN ACTION

Learning how to properly decipher and forge citations to primary sources is part of the most basic training of any classicist. When building an expert system which extracts automatically such citations from unstructured texts, we need some sort of surrogate which provides us with such domain knowledge. In this section we shall examine how a KB based on HuCit can be automatically populated and queried in order to support this information extraction task.

5.1 Querying the Knowledge Base

Let us consider now what are the types of queries that can be ran against our KB. As an example let us take the one provided by Crane in [3] to exemplify the challenges of capturing canonical citations and their meaning:

> Citation identification is a particular case of named entity identification that focuses on recognizing particular: e.g., determining whether the string "Th. 1.33" refers to book 1, chapter 33 of Thucydides, line 33 of the first Idyll of Theocritus or something else?

Although 'Th.' is not a common way of abbreviating the name of Thucydides, what we have here is a typical example of ambiguous citation. In fact, the string 'Th.', when considered without its surrounding context, can be a reference to the Greek historian Thucydides as well as to Theocritus. The idea underlying HuCit is that information about the structure of canonical texts can be exploited to improve the accuracy of an automatic system in determining the correct meaning of a citation in cases like the one above.

A first type of query can help us validating citations, that is checking if whether the scope of the citation exists or not (in the example above, "1.33" is considered the scope of the citation). Following the example of "Th. 1.33", a query against our KB will look for works whose title can be abbreviated as "Th." and whose canonical text structure has a first-level element "1" and a second-level element "33". Such a query will return two matches: book 1, chapter 33 of Thucydides' *Histories* and Idyll 1, line 33 of Theocritus' *Idylls*. This kind of query is not enough, in this case, to rule out all other possibilities, thus leaving only one valid candidate.

In order to refine the results of the query above, we may try to search for clue words within the context of the citation. These are words such as 'book', 'chapter', 'line' or 'idyll' and tell us something useful about the text that is being cited. In Thucydides' case, its `CanonicalTextStructure` will contain instances of `TextElement` with type 'Book', 'Chapter' or 'Section', whereas in Theocritus' case the `TextElements` will be of type 'Idyll' or 'line/verse'. At this point, one can try to match words from the context against such type labels and eventually narrow down further the list of candidates.

5.2 Citations as Annotations: HuCit and OAC

The Open Annotation Collaboration (OAC) ontology is a rather simple yet very flexible semantic format which is suitable to publish not only user-generated annotations, but also annotations that are produced automatically, such as for example linguistic annotations and specifically annotations of named entities. It is simples as it represents any annotation as an object having a body (`oa:Body`) and a target (`oa:Target`): the target is the portion of an object, not necessarily text, that is annotated, whereas the body is the actual content of the annotation. Additionally, an annotation can have basic metadata such as creator, creation date, etc. Its flexibility lies in the fact that there are virtually no restrictions on the object type of the body or the target.

As they can be referred to by means of URIs, the object type (i.e. its class) is determined when the URI is resoved and can be any, as long as its semantics are defined in a machine-readable format (e.g. OWL/RDF).

In the latest version of the OAC specifications $(1.0)^{16}$ the new class `oa:Motivation` was introduced in order to provide some information about the reason why a a given annotation was created. For example, an annotation can be created to reply to another annotation. OAC defines already a number of instances of this class such as `oa:highlighting`, `oa:replying` or `oa:identifying` that can be used as values for the `oa:motivatedBy` property. Among these pre-defined instances of `oa:Motivation` one seems particularly suitable to represent named entity annotation, and therefore also the annotation of canonical citations, that is `oa:identifying`, which

> [...] represents the assignment of an identity to the target resource(s). For example, annotating the name of a city in a string of text with the URI that identifies it.

HuCit can be combined with OAC to store and publish the annotations produced by a service which extracts canonical citations from text: the motivation behind such annotations can be expressed by means of `oa:identifying`; `oa:Target` is the portion of a text that was tagged as being a canonical citation, whereas `oa:Body` identifies its content, that is in HuCit terms one or more `TextElements` of the `CanonicalTextStructure` of a given work.

6. CONCLUSIONS AND FURTHER WORK

The work we presented in this paper is a first attempt to provide a formal modeling of citations to primary sources in Classics.

The benefit of the presented approach–reifying the content of citations, rather than treating them as mere relationships between documents–is two-fold: 1) since the content of a citation becomes an instance on its own, it can be given a URI and can link to related resources on the (Semantic) Web; 2) the URI pointing to the content of a citation can be used as value of the `oa:hasBody` property, thus allowing us to model in depth our domain of interest while maintaining some degree of interoperability with other resources and services that make use of the OAC standard to describe annotations.

Another research direction that might be considered in the future is exploring to what extent the intention behind citing primary sources in Classics can be determined and consequently modeled in our ontology, in a way that is similar to what CiTO already does with its taxonomy of citation types. The main problem with this that is that the usual way of citing sources, for example as support to an argument or claim (e.g. "vd.", "cfr.", "cf. e.g."), can be so vague or underspecified that is often hard, if not impossible, to establish the purpose of an author in citing a given text [4].

---

^{16}http://www.openannotation.org/spec/core/

7. REFERENCES

[1] J. D. Bolter. Hypertext and the Classical Commentary. In *Accessing antiquity : the computerization of classical studies*, pages 157–171. University of Arizona Press, Tucson, 1993.

[2] P. Ciccarese, E. Wu, G. Wong, M. Ocana, J. Kinoshita, A. Ruttenberg, and T. Clark. The SWAN biomedical discourse ontology. *Journal of Biomedical Informatics*, 41(5):739–751, 2008.

[3] G. Crane, B. Seales, and M. Terras. Cyberinfrastructure for Classical Philology. *Digital Humanities Quarterly*, 3(1), 2009.

[4] D. Fowler. Criticism as commentary and commentary as criticism in the age of electronic media. In G. W. Most and D. Fowler, editors, *Commentaries = Kommentare*, number 4 in Aporemata: Kritische Studien zur Philologiegeschichte. Vandenhoeck & Ruprecht, Göttingen, 1999.

[5] T. Heath and C. Bizer. *Linked Data: Evolving the Web into a Global Data Space*. Synthesis Lectures on the Semantic Web. Morgan {&} Claypool Publishers, 2011.

[6] R. Mizoguchi. Tutorial on Ontological Engineering Part 3: Advanced Course of Ontological Engineering, 2004.

[7] M. Pasin and R. Mizoguchi. Moving EMLoT towards the web of data: an approach to the representation of humanities citations based on role theory and formal ontology (forthcoming), 2013.

[8] L. D. Reynolds and N. G. Wilson. *Scribes and Scholars: A Guide to the Transmission of Greek and Latin Literature*. Oxford University Press, USA, 3 edition, 1991.

[9] M. Romanello. Creating an Annotated Corpus for Extracting Canonical Citations from Classics-Related Texts by Using Active Annotation. In A. Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing. 14th International Conference, CICLing 2013, Samos, Greece, March 24-30, 2013, Proceedings, Part I*, volume 1 of *Lecture Notes in Computer Science / Theoretical Computer Science and General Issues*, pages 60–76. Springer Berlin Heidelberg, 2013.

[10] M. Romanello, M. Berti, F. Boschetti, A. Babeu, and G. Crane. Rethinking Critical Editions of Fragmentary Texts By Ontologies. pages 155–174, Milano, Italy, 2009.

[11] M. Romanello, F. Boschetti, and G. Crane. Citations in the digital library of classics: extracting canonical references by using conditional random fields. In *Proceedings of the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries*, NLPIR4DL '09, pages 80–87, Morristown, NJ, USA, 2009. Association for Computational Linguistics.

[12] D. Shotton. CiTO, the Citation Typing Ontology. *Journal of biomedical semantics*, 1(Suppl 1):S6, 2010.

[13] R. Simon, E. Barker, and L. Isaksen. Exploring Pelagios: a visual browser for geo-tagged datasets. In *International Workshop on Supporting Users' Exploration of Digital Libraries*, 2012.

[14] N. Smith. Digital Infrastructure and the Homer Multitext Project. In G. Bodard and S. Mahony, editors, *Digital Research in the Study of Classical*

Antiquity, pages 121–137. Ashgate Publishing, Burlington, VT, 2010.