Reflections on Cultural Heritage and Digital Humanities: Modelling in Practice and Theory

Arianna Ciula
University of Roehampton
UK
arianna.ciula@roehampton.ac.uk

Øyvind Eide
Universität Passau
Germany
oyvind.eide@uni-passau.de

ABSTRACT

Computer based modelling in cultural heritage has focused on database development, generalised as data standards and, since the 1990s, also formal ontologies. Modelling in digital humanities has had its core in textual scholarship, including close reading and text encoding of literary and historical sources as well as models of text corpora, usually relying on statistical methods. Integration between the two modelling paradigms has been undertaken at the practical level. This paper goes beyond pragmatic concerns by focusing on comparing the two modelling traditions at a more abstract level.

To this end, one core standard development undertaken in each domain is selected: CIDOC's Conceptual Reference Model (CRM) for modelling in cultural heritage—narrowed down to museum documentation—and Text Encoding Initiative (TEI) for modelling in digital humanities—narrowed down to textual scholarship. This does not imply that these two standards are only used in the two areas mentioned above, rather that their main focus has been in those areas. We will use the two standards to investigate what is meant by modelling in the two communities, thus, clarify the differences and the similarities between the concepts of modelling and models as they are used in each community. Partly this will be done based on a survey of previous literature, and partly by an investigation into modelling practices within these two standardisation initiatives. Minutes and reports, descriptions of the standards themselves and their developments, mailing lists threads, and the participants' own experience in the development of the respective standards have informed this study.

Categories and Subject Descriptors

I.6.4 [**Simulation and Modelling**]: Model Validation and Analysis; I.6.5 [**Simulation and Modelling**]: Model Development

DATeCH 2014, May 19 - 20 2014, Madrid, Spain

ACM 978-1-4503-2588-2/14/05 ...$15.00.
http://dx.doi.org/10.1145/2595188.2595207.

Keywords

Digital Humanities, Cultural Heritage, Modelling, Museum Informatics, Text Encoding, CIDOC-CRM, TEI

1. INTRODUCTION

Different modelling languages and theories are used in the humanities and in cultural heritage. A long term aim of our research is to examine a number of these languages in order to understand how they are created and used. For the time being we focus on one of the central modelling languages for each area, being aware that despite some overlaps, the scopes of modelling in the two standards are categorically different. CIDOC-CRM aims at describing real world objects as they are represented in museum information systems, while TEI is mainly used to represent textual features without any assumptions about their reference functions.

The TEI[1] or Text Encoding for Interchange was born as a research project in 1987 and evolved into a Consortium structure which was established in 2001. It is a community based effort to produce a set of Guidelines to encode any type of text in digital form.[2] It currently relies on XML as its manifestation, but it could potentially be expressed in other formalisms. The development and maintenance of the Guidelines is taken care of by the technical council, an elected group of around 10 individuals who meet face to face once or twice a year. Their work is supported by specific chartered groups of external experts, including Special Interest Groups (SIGs). Currently most of the work on the guidelines happens remotely through the support of tracking systems and versioning platforms. However, face to face meetings are still crucial when decisions on the actual modelling of the TEI scheme[3] take place. The TEI development is anchored to traditional humanities, especially textual scholarship; however, the community of contributors to the development of the standard is rather hybrid in terms of interests and discipline affiliation. It includes among others computer scientists, linguists, philologists, lexicographers, librarians, classicists, as well as experts in literary studies, manuscript studies, and information studies.

[1]Webpage: http://www.tei-c.org/ (all URLs were checked January 17, 2013)
[2]One part of the TEI scheme, namely, the model for feature structures, is an ISO standard (ISO 24610-1:2006 and ISO 24610-2:2011) [1, 16].
[3]In this article we use the term 'TEI scheme' as synonymous for TEI overall model or conceptual structure of the TEI standard as a whole - what the Guidelines refer to as "TEI Abstract Model" [25, sec. 23.4.3].

CIDOC-CRM[4] is a formal ontology[5] intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that cultural heritage information can be mapped to. This provides the "semantic glue" needed to mediate between different resources, such as those published by museums, libraries and archives. CIDOC-CRM has been an ISO standard (ISO 21127:2006) since 2006. The work is coordinated by the CIDOC CRM SIG, a working group of CIDOC.[6] The standard is maintained partly by mailing list discussions and publicly available lists of issues at the webpage, while most decisions are made during face to face SIG meetings taking place 2–3 times a year. The participants come from museums, other cultural heritage institutions, universities and research centres, and private companies, mainly SMEs.

2. MODELLING IN THEORY

Generally, when the term 'model' is used in computing, it recalls the strict sense of 'data model'. Data models as instantiations of the TEI and of the CIDOC-CRM standards can of course be created. For example a specific TEI-conformant XML schema is a data model in this sense. However, the use of this term is ambiguous, spanning from the lower level meaning of instances, realisations, or implementations of a general conceptual structure (e.g. a TEI XML manuscript catalogue based on the TEI recommendations on how to encode a manuscript description) to the broader meaning of a conceptual model (e.g. the *Ordered Hierarchy of Content Objects* or OHCO model to represent a hierarchical organisation of texts, a model dominant in XML instances and therefore also in TEI). Whatever meaning one adopts, our attempt here is to examine the process of model creation or modelling rather than focusing on static or abstract data models *per se*.

To grasp the processual nature of models and move the focus from data models to the creation of models as standards themselves or as instantiations of standards, we ground our analysis on a general concept of modelling and model as recently defined in [15]. The definition covers disparate uses of the term 'model', such as modelling in art, scale models in design processes, mathematical equations, and photographs. Models are understood within a subject- and context-dependent semiotic structure as following:

> [...] models can be regarded as a specific kind of signs and, more precisely, as icons, i.e. as signs which are characterized by a (subjectively recognized) similarity relation between sign and object, i.e. between model and original.
>
> [...] models are determined by a semiotic structure in which a subject intentionally uses an object, the model, as a sign for another object, the original, in the context of a chosen theory or language in order to attain a specific end by instituting a representational relation in which the syntactic structure of the model, its attributes and relations, represents by way of a mapping the properties of the original, which hence are regarded as or are postulated to be similar in a relevant manner [15, 3419].

The concept of similarity between models and objects to be modelled, as described in [15], allows us to capture both the processes usually referred to as 'mapping'—performed by CIDOC-CRM users when fitting specific museum documentation resources into the CIDOC-CRM conceptual model—and as 'document analysis'—performed by TEI users when categorising specific textual features to be annotated in compliance with the TEI scheme.[7] This definition also accounts for the dependence of any act of modelling on its theoretical setting and use of language. Hence, in turn, the models so created—e.g. a database of a museum collection modelled in CIDOC-CRM or a critical edition modelled in TEI—represent the semantics of their respective standards—CIDOC-CRM conceptual model and TEI scheme respectively; in so doing they also reflect certain corresponding theoretical approaches to museum documentation and textual studies.

The powerful epistemic values of models reside in the fact that while being dependent on theory, models do transcend them. This is recognised in [15]. Linking this to modelling as understood in digital humanities, in particular [17], we could say that by practice—through cognitively rich modelling processes—models make their creators re-think and question their theories. Modelling is dynamic and heuristic.

Within a similarly generalised idea of modelling one can distinguish between two different, yet overlapping, aims modelling can have: modelling for and modelling of [17]. Like a prototype in industrial design, a model for something is part of the process of making something new, e.g. to test an hypothesis or to try out specific functionalities. A model of something is an abstraction created on the basis of but not stopping at a representational relationship with something already existing or thought to have existed. 3D reconstructions in archaeology are models in this sense.

Both TEI and CIDOC-CRM are used for 'modelling for'. A common way of using TEI is to make a TEI encoding of a text for the purpose of producing a printed or digital edition. CIDOC-CRM mappings of museum databases are often created with the purpose of data integration. CIDOC-CRM concepts are "instantiated as sets of statements that provide a model of reality" [4, iv], and the integration potential follows from the fact that the systems to be integrated are talking about the same real world.[8] While these activities will generate new knowledge and can very well be part of research projects, they still have as their main practical goal to create a new information object: the encoded text and the mapping, respectively. However, when the making of a TEI encoding and a CIDOC-CRM based model is part of a research strategy with the aim to question the theory

[4]Webpage: http://www.cidoc-crm.org/

[5]"In the context of computer and information sciences, an ontology defines a set of representational primitives with which to model a domain of knowledge or discourse. [...] Ontologies are typically specified in languages that allow abstraction away from data structures and implementation strategies; in practice, the languages of ontologies are closer in expressive power to first-order logic than languages used to model databases. For this reason, ontologies are said to be at the 'semantic' level, whereas database schema are models of data at the 'logical' or 'physical' level" [12].

[6]CIDOC is the International Committee for Documentation of the International Council of Museums (ICOM).

[7]See the section of the TEI guidelines where the notion of conformance is discussed extensively [25, sec. 23.4].

[8]In TEI modelling of texts, the instances can include fictitious as well as real worlds.

underlying the modelling process rather than concentrate exclusively on the production of something new, 'modelling for' gains a self-reflective perspective; it becomes 'modelling of'.

As it is evident, this is not a clear cut distinction – modelling for will also involve modelling of and vice versa [16, 372]. But the former is more geared towards implementation of the unknown, whereas the latter is more geared towards (dynamic) representation of what is known. In digital humanities we do not only create models as fixed structures of knowledge, but also as a way to investigate a series of temporary states in a process of coming to know. The point of this kind of a modelling exercise lies in the process, not in the model as a product. Eventually, when the model is perceived no more as expanding the knowledge about the objects it models, but restrictive and limiting [7], an epistemic shift is made: one model is superseded and a new one envisaged. In experimental and design settings, this is what is normally referred to as iterative cycle (e.g. [18]).

The processes of modelling mentioned above are based on the use of already existing standards, they are instantiations of the standard. What about the creation of the standards themselves? It is clear that the main mission of both TEI and CIDOC-CRM is to create models for. The standards were conceived in order to be used as tools for text and information encoding, exchange, publication, etc. However, both standards are also created as models of something. For instance, TEI includes a model of what the main constituent parts of a theatre play might be. While that model can be used in the encoding of plays, it is also a model of a large number of existing plays known to the creators of the standard. Similarly, when CIDOC-CRM includes a model of what happens when an object is on loan from one museum to another, it is a model for creating mappings from existing information systems. It is also a model based on how this process is implemented in a number of different museum information systems, and, through that, of how museums operate.

3. PREVIOUS COMPARISONS BETWEEN TEI AND CIDOC-CRM

The TEI Ontologies SIG–officially established at the annual TEI meeting in Baltimore in 2004—has been open to any work connecting TEI to external ontologies, but its focus from the outset was on CIDOC-CRM. In this section we will first give a short outline of the history behind the establishment of the SIG, and then cover the development since 2004 along two lines: practical interlinking between the two standards at the level of data sets and theoretical works studying the relationship between the two standards at class level, that is, comparing the elements and how they are structured in TEI to the entities and properties in CIDOC-CRM.

In the early 1990s the national Norwegian Documentation Project was given the task of developing a database for the archaeological collections held by four major Norwegian museums. The traditional method of reading through source texts and enter the information which seemed to be relevant in a normalised form was not considered as adequate by the project team, as the link to the original text was lost, thus leading to a loss in scholarly reproducibility. Hence, a choice was made to encode museum catalogues in SGML, with detailed semantic markup, and then extract relevant information into the databases being established, keeping the links back to the source SGML documents [14].

While this system did not use neither TEI nor CIDOC-CRM directly,[9] it established the basic idea of using SGML and later XML encoding of texts describing museum objects as a way of connecting textual source information directly to museum databases. Working with nineteenth century museum catalogues highlights the need for reading the textual information in light of the concepts of the time, but in principle all texts express a world view which may not coincide neither with the contemporary reader's world view nor with the assumptions underlying a specific information system. The main contribution from the text encoding community to culture heritage information systems was the basic understanding of texts, also seemingly neutral texts describing the real world, as culturally situated.[10]

Several projects using TEI encoding took up the use of CIDOC-CRM in the years following the establishment of the SIG. One example is the Fine Rolls of Henry III project [5], where some features of the TEI markup (e.g. from element content to relationships based on implicit nesting of elements) were extracted—mainly using XSLT—to populate an ontology built around CIDOC-CRM entities and properties. As the markup of the text was further developed, a new import from TEI to RDF/OWL would map new data to the ontology. The main objective of this effort was to accompany a detailed structural and syntactic encoding of the text sources with relational information, e.g. historical information mainly about persons, places and subjects. While TEI XML encoding would mainly operate at the level of linearity of the text itself, CIDOC-CRM and other ontological modelling was used to overcome such linearity while still anchoring specific statements to the text itself. A strong theoretical framework for an approach combining and synchronising markup with a "contextual and procedural information" system as a viable practical method to account for and process textual mobility was later provided by [3].

The CLAROS project developed a fully fledged system for establishing CIDOC-CRM compliant RDF triples based on TEI encoded information. The CLAROS team created and implemented a workflow where first the relationship of TEI elements to known CIDOC CRM concepts were established in a formal way, maintained in a single document with the mapping guidelines, and then developed actual mapping code from TEI XML to RDF XML [21].

[19] is one of the outcomes of comparisons between the two standards at type level. Differences between the standards were studied by "comparing the expressive power of the real world descriptions TEI P5 by mapping central parts of the CIDOC CRM onto TEI P5" [19, 161]. This was done by comparing the ontological elements in TEI P5 with CIDOC-CRM, using the latter as a yardstick against which the former was evaluated.[11]

[9]Besides basic structural elements, TEI had little to offer for these text types at the time. See [13] for an attempt to integrate the code system used in The Documentation Project with an early version of the TEI ODD formalism.

[10]This observation offers an alternative historical perspective to the current critique of digital humanities as the revenge of positivism in the humanities (see e.g. [10]).

[11]The main goal of the article was to point out what was needed to make sure that TEI was part of the ongoing harmonisation process between important standards in cultural heritage information. Results from the article were later

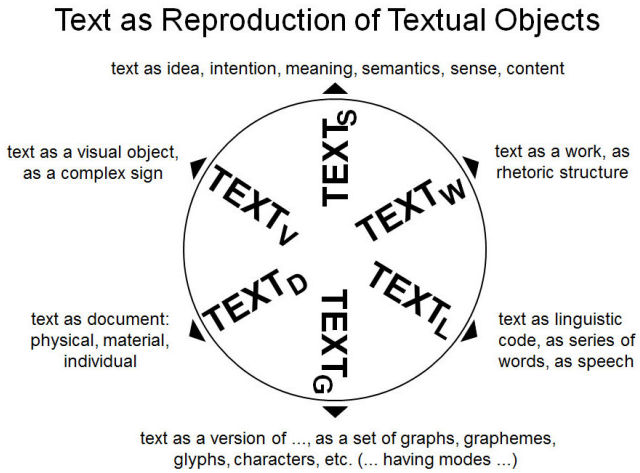

Figure 1: Sahle's pluralistic model of text [23], first presented at Digital Humanities 2006.

This article goes beyond the reviewed work in that it not only compares the expressive power of the two standards in specific areas, such as the modelling of places, persons, or events, but will attempt at comparing the standards themselves. Such a comparison at the level of standards includes studying what types of information each of the standards can express and how this information is modelled in relation to specific selected features of the objects being modelled and specific formalisations being adopted.[12] Note that this is an early report of work in progress; many questions are left open.

4. MODELLING IN PRACTICE

Free text as we find it in existing paper-based or digital editions cannot be considered as unstructured in any strict sense, since any kind of text embeds conventions, from punctuation to bibliographic features. Nevertheless, there are still significant differences between a table or relation based structuring of data one finds in most museum systems and the richness and many-facetedness found in, say, a novel. This explains much if not all of the significant differences between the two standards: CIDOC-CRM was created in order to formalise the former, TEI was created to encode also the latter and its potentially multidimensional textual features.[13] Sahle has visualised this expansive aim of the TEI scheme in the wheel chart reproduced as figure 2, which is based on his text model reproduced as figure 1.

Besides the main purpose of each standard, historical contexts of development are also relevant to understand how they have mutated. The documentation form of the TEI Guidelines, for instance, has evolved from a format inspired by the structure of books into a more modular system.[14]

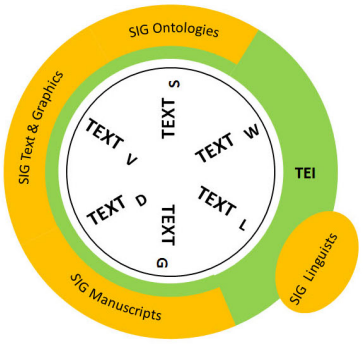

Figure 2: Sahle's mapping of the TEI model onto figure 1 [23], cf. [24, part 3 p. 371–390].

However, despite the rigour of its formal expression (currently represented by ODD [2], the TEI XML source format standing for *One Document Does it all*) and the rendering of element specifications in tabular form, the semantics of the standard resides mainly in the narrative as it unfolds in each chapter of the Guidelines.

The standard contains around 500 elements—i.e. types of textual components and annotation labels—and accounts for many different ways to model the 'same' textual phenomena or object. While the latter has been described as a point of strength and flexibility in accommodating different encoding purposes and theories, it is also criticised as a main drawback with respect to interchange across resources. It is not by chance that communities focused on specific kind of texts have attempted at breaking down the complexity of TEI models into more restricted but semantically explicit structures. An example of the latter is the concept of 'crystals' proposed in [22] to account for the organisation and semantics of lexical structures via constraints of hierarchical dependency or semantic inheritance.

The CIDOC-CRM is "a formal ontology intended to facilitate the integration, mediation and interchange of heterogeneous cultural heritage information." [4, i]. While this wording in itself could give the impression it is created to cater for all cultural heritage information including free text, the standard as a whole, and the history behind it, clearly shows that the main target of representation is information as recorded in museum information systems. While the latter includes a variety of pre-digital systems, such as index cards, protocols, and acquisition catalogues, museum documentation is read differently from the text types most commonly encoded in TEI, such as poetry, drama, or novels.

As outlined in the introduction to the standard, CIDOC-CRM is developed in a regime of strict design principles [4, xv–xviii], see also [6]. For instance, the ontology makes a clear distinction between the standard itself and compliant document instances as valid and well formed on the one hand, and conclusions about the truth on the other. The latter is the task of open-ended scholarly hypothesis building

used for further development of the TEI scheme [26].

[12]Some remarks in this direction were made already in the 2009 article, such as pointing out the difference between CIDOC-CRM as a formal ontology and TEI as a less formal set of recommendations [19, 167].

[13]This is discussed in [20].

[14]For a full account see [1]. Note that embedding such developments in a specific historical context has not only sociological implications that go beyond the scope of this paper, but also direct consequences on the settings where modelling strategies (with respect to the refinement of the standard) take place. The organisational structures has developed from a research project answerable to a self-selecting group of experts through a centralised structure in which all the work is done by a small number of people under the control of a small executive authority to a more distributed structure in the present decade.

systems. Thus, when a new fact (i.e. a statement that a painting was painted by A and not B) is added to a CIDOC-CRM model it may very well be conflicting with another factual statement already in the system. Such conflicts are not contradictions in the strict sense. They can be compared to a that-sentence in natural languages [11]; for example, "Peter said that it was raining." The sentence can be true even if it was not raining, as long as the fact that Peter made that utterance is true.

Given their purposes and origins, it is no surprise that CIDOC-CRM and TEI are two quite different modelling standards made accessible in respectively different documents. What meets the eye is, on the one hand, a rather concise book describing the CIDOC-CRM ontology in great detail, with its use not being exemplified much in the standard itself. TEI, on the other hand, is an extensive set of Guidelines available as highly cross-referenced hypertext (in HTML, PDF, EPUB and MOBI formats) and printed volumes, where the formal definition of each element represents only one access point to the standard. The latter can be better understood through the lengthy descriptions of use cases.[15] The rendering of these two standards into documents are similar in that they are both practice based and they both make theoretical assumptions about the domain they model; however the way they present such assumptions and the components of the standards varies significantly.

In CIDOC-CRM all entities are presented in a standardised form consisting of inheritance (super- and subclasses), scope notes, examples, and properties available. Property presentations consist of domain, range, qualifications, scope notes, and examples. Visualisations of the standard as a whole are available as interactive systems. In addition, two graphical presentations of parts of the model are included in the standard document: one for reasoning about spatial information and one for reasoning about temporal information. Furthermore, the heritage structures for entities and properties are presented as indented lists.

The presentation of the TEI standard follows a very different pattern. The Guidelines are divided into 23 chapters following a standardised organisation including the following components: thematic overview; a summery table of new elements with descriptions; usage examples, usually based on real sources; links to the reference specifications for the elements and for any model or attribute class mentioned.

The 'reference specification' view or access to the Guidelines, as mentioned above, is presented in tabular form and consists of:

- a label or name element accompanied by a brief description (crucial to disambiguate between the natural language use of the English relevant term and the specific TEI connotations);
- module/s (corresponding to one or more chapters where that element is discussed in prose);
- attributes (listing attribute classes and their extension);
- model class (affiliation to grouping of elements based on content);
- elements that can contain the element being described;
- elements that the element being describe can contain;
- formal declaration (e.g. expressed in RNG compact syntax);
- one or more selected examples of usage.

To exemplify these differences, we will study in some detail how the concept of place name is described in each of the two standards. In CIDOC-CRM, the relevant entity is E48 Place Name, a subclass of E44 Place Appellation. The scope notes for E48 Place Name reads:

> This class comprises particular and common forms of E44 Place Appellation.
>
> Place Names may change their application over time: the name of an E53 Place may change, and a name may be reused for a different E53 Place. Instances of E48 Place Name are typically subject to place name gazetteers [4, 20].

It also includes 4 examples: "Greece", "Athens", "Geneva", and "LacLéman". When used, CIDOC-CRM E48 is connected to other entities through properties, of which the most important are:

- P139 *has alternative form* connects two instantiations of E48 Place Name. It is used for alternative spellings and for transliterations.
- P87 *identifies* connects a E48 Place Name to the E53 Place it identifies.

In addition to this, E13 Attribute Assignment events can be used to connect names to places through explicit naming events, where time, place, and the responsible actor can be recorded. We will not go into the details of attribute assignment events here.

The relevant TEI chapter where <placeName> is discussed [25, sec. 13.2.3] exemplify how it can be used to make explicit information about the place it refers to (geopolitical structures as expressed in the naming itself), to link to other annotations with respect to the place it refers to (as expressed elsewhere from the text snippet being encoded e.g. in a gazetteer) or with respect to the name itself (toponomastic features). Such variety creates ambiguity but also reflects the transparency of natural language in mapping a name to its named entity.[16] Here follow two XML examples selected freely from chapter 13 and the relevant reference specification in the Guidelines:

```
<placeName>
<settlement>Rochester</settlement>
<region>New York</region>
</placeName>

<placeName ref="tag:projectname.org,2012:NY1">
New York</placeName>
```
[17]

[15] No visualisation of the TEI scheme as a whole or as single components are currently included in the standard.

[16] See the debate on the TEI-L mailing list about the semantic ambiguity brought in by the use of the attribute @type within <placeName> as potentially referring to the name being encoded or to the place that name refers to [9].

[17] Note that the identification and possibly information about the place this name refers to are stored elsewhere.

The primary role of these comparable concepts differ between the two models. Whereas in CIDOC-CRM the modeller establishes connections between place names and other entities, in TEI the linking to other elements is only one of the possible reasonings around that element usage. TEI <placeName> can also be used to encode names of fictitious places as exemplified in the Guidelines. In CIDOC-CRM, the name of a fictitious place is not an E48 Place Name.[18]

5. CONCLUSIONS

Despite significant differences, the two standard documents share a similar core of presentational structure and functionalities. Both describe general features of their underlying standards meant to model particulars or instances. Neither attempt at providing names for classes encapsulating a definition. Names are rather used as labels; scope or specification notes and examples encircle the specific target of each type. Thus, they include general descriptions in natural language. But they also include examples as a way to show how particulars are the real targets of the type or class descriptions (i.e., of TEI elements and CIDOC CRM entities).

The example in the previous section confirms, however, a fundamental difference in what is seen as context. TEI encoding appears in a textual context—generally made of words—intermixed with XML tags. Even by abstracting away the text itself,[19] the stripped out XML tree constitutes basically a model of one way of seeing the text structure: the place name is part of a sentence which is part of a paragraph which is part of a chapter and so on.[20] In the CIDOC-CRM model, on the other hand, the place name is related to the role it has in a system documenting statements about the world. It is primarily and foremost connected to the place it functions as a name for.

As we saw, this relational aspect can also be expressed in TEI. If one has a representation of places as objects in a real or fictitious world, either as part of the TEI header or in some other information system, a two way link between each place name occurrence in the text and the place representation can be made with the use of an XML ID/IDREF connection. But this is different from CIDOC-CRM in several respects. When a second layer —graph-like rather than tree-like component— is added to the XML document, the model gains in expressive power but loses in manipulative power, e.g. with respect to XML derivative technologies used to process it. Secondly, the place name in the encoded text is a particular use (occurrence) of that name, a different concept compared to name(s) nested, for instance, in the place element in the header and representing selected identifiers for that place. The latter scenario—place names elements nested in place elements in the TEI header or separate TEI document—is in line with the use of place name entities in the CIDOC-CRM, where the names are identifiers of a place. Hence, while the TEI scheme in its creation is also based on an abstraction process that transcends particulars, its use in specific encodings is intrinsically dependent on and anchored to the objects (texts) being modelled. While CIDOC-CRM is based on a specified model of the world, no such model is explicit in TEI, which encompasses many models, some more explicit than others. This was also noted in 2009:

> The new and revised modules described in TEI P5, Chapter 13 *Names, Dates, People, and Places*, are defined without any explicit references to any specific ontology and are designed to cover a wide variety of real world descriptions. Here the authors of the guidelines apparently intend to follow an ontological neutrality ideal. It is however not possible to define this set of elements without having (several) implicit conceptual models in mind. This part of the TEI P5 may have been clearer if these mental models had been harmonized and made explicit in the guidelines." [19]

In this paper we have not only compared what objects the two standards are meant to model, but also how the standards themselves are modelled and made accessible to users, as a step towards understanding how the standards as complete systems steer, in turn, the modelling processes.

We have exemplified how modelling languages in the cultural heritage and digital humanities domains are both enablers for and obstructors against expressing certain knowledge and understanding of their respective target systems. While this exercise is interesting in itself as an investigation into modelling strategies, it also has a more pragmatic aim of raising our own awareness about the choices that are made in certain modelling practices. Rather than being perceived as a divider between communities and traditions, such awareness enables a certain freedom. The different approaches combined can help envisaging imaginary constructs which can be used to model cultural artefacts and their interpretations in new ways.

6. ACKNOWLEDGEMENTS

The authors would like to thank Christian-Emil Ore for his comments and contribution in finalising this article.

7. REFERENCES

[1] L. Burnard. The Evolution of the Text Encoding Initiative: From Research Project to Research Infrastructure. Journal of the Text Encoding Initiative, (5), 2013.

[2] L. Burnard and S. Rahtz. RelaxNG with Son of ODD. In Proceedings of Extreme Markup Languages 2004, 2004.

[3] D. Buzzetti and M. Rehbein. Towards a model for dynamic text editions. In Proceedings from Digital Humanities 2008, pages 78–81, Oulu, 2008. University of Oulu.

[4] CIDOC. Definition of the CIDOC Conceptual Reference Model. CIDOC, [Heraklion], 2011. Produced by the ICOM/CIDOC Documentation Standards Group, continued by the CIDOC CRM Special Interest Group. Version 5.0.4, December 2011.

[18] For a discussion of the role of fictitious objects, see [8] and further replies.

[19] We will not enter the debate of what text is here, just note that even if we strip away the markup from a TEI encoded file, what is left is not 'the text' in a pure form. It is rather 'a text' with other conventions, annotations and interpretations not represented using the TEI scheme embedded, e.g. as punctuation marks.

[20] There are of course other possible models of the same text. See [25, ch. 20] on the problem of overlapping text structures.

[5] A. Ciula, P. Spence, and J. M. Vieira. Expressing complex associations in medieval historical documents: the Henry III Fine Rolls Project. Literary and Linguistic Computing, 23(3):311–325, 2008.

[6] M. Doerr. The CIDOC Conceptual Reference Module[!]: An Ontological Approach to Semantic Interoperability of Metadata. AI magazine, 24(3), 2003.

[7] Ø. Eide. Ontologies, data modelling, and TEI. In F. Ciotti and A. Ciula, editors, The Linked TEI: Text Encoding in the Web. Abstracts of the TEI Conference and Members Meeting 2013. Proceedings, pages 26–30, Rome, 2013. DIGILAB Sapienza University & TEI Consortium.

[8] Ø. Eide. Re: disguises? TEI-L, 2014. Sun, 23 Feb 10:51:15 +0100. http://listserv.brown.edu/archives/cgi-bin/wa?A2=ind1402&L=TEI-L&F=&S=&P=131545.

[9] T. Elliott. placeName types. TEI-L, 2013. Fri, 17 May 13:10:18 -0500. http://listserv.brown.edu/archives/cgi-bin/wa?A2=ind1305&L=tei-l&D=0&P=61357.

[10] T. Eyers. The Perils of the "Digital Humanities": New Positivisms and the Fate of Literary Theory. Postmodern Culture, 23(2), 2013. DOI: 10.1353/pmc.2013.0038.

[11] G. Frege. Über Sinn und Bedeutung. Zeitschrift für Philosophie und philosophische Kritik, vormals Fichte-Ulricische Zeitschrift, NF 100:25–50, 1892.

[12] T. Gruber. Ontology. In L. Liu and M. T. Özsu, editors, Encyclopedia of Database Systems, pages 1963–1965. Springer US, [S.n.], 2009.

[13] J. Holmen and Ø. Eide. Reading Gray Literature as Texts. Semantic Mark-up of Museum Acquisition Catalogues. In Proceedings from CIDOC 2006, 2006. http://www.edd.uio.no/artiklar/teknikk_informatikk/CIDOC2006/EIDE_HOLMEN_Reading_Gray_Literature.pdf.

[14] J. Holmen, C.-E. Ore, and Ø. Eide. Documenting two histories at once: Digging into archaeology. BAR International Series, 1227:221–224, 2004.

[15] B. Kralemann and C. Lattmann. Models as icons: modeling models in the semiotic framework of Peirce's theory of signs. Synthese, 190(16):3397–3420, 2013.

[16] B. Mahr. Information science and the logic of models. Software & Systems Modeling, 8(3):365–383, 2009.

[17] W. McCarty. Humanities computing. Palgrave Macmillan, Basingstoke, 2005.

[18] J. Nielsen. Iterative User Interface Design. IEEE Computer, 26(11):32–41, 1993.

[19] C.-E. S. Ore and Ø. Eide. TEI and cultural heritage ontologies: Exchange of information? Literary & Linguistic Computing, 24(2):161–172, 2009.

[20] E. Pierazzo. Conceptual model of Text, Documents and Work—Part 1. Blogpost from August 6, 2013. URL: http://epierazzo.blogspot.co.uk/.

[21] S. Rahtz. TEI to CRM. In TEI Ontologies SIG meeting, Zadar, 2010. http://tei.oucs.ox.ac.uk/Oxford/2010-11-12-CRM/talk.pdf.

[22] L. Romary and W. Wegstein. Consistent Modeling of Heterogeneous Lexical Structures. Journal of the Text Encoding Initiative, (3), 2012.

[23] P. Sahle. Modeling Transcription. Knowledge Organization and Data Modeling in the Humanities: An ongoing conversation. Workshop at Brown University March 2012. URL: http://datasymposium.wordpress.com.

[24] P. Sahle. Digitale Editionsformen, Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Books on Demand, Norderstedt, 2013.

[25] TEI Consortium. TEI P5: Guidelines for Electronic Text Encoding and Interchange. [2.5.0]. [July 26 2013]. TEI Consortium, [S.n.], 2013.

[26] C. Wittern, A. Ciula, and C. Tuohy. The making of TEI P5. Literary and Linguistic Computing, 24(3):281–96, 2009.