

# ***Prosopography and Computer Ontologies: towards a formal representation of the 'factoid' model by means of CIDOC-CRM***

John Bradley, Michele Pasin  
Department of Digital Humanities  
Kings College, London  
{john.bradley, Michele.pasin} @kcl.ac.uk

Structured Prosopography provides a formal model for representing prosopography: a branch of historical research that traditionally has focused on the identification of people that appear in historical sources. Pre-digital print prosopographies, such as Martindale 1992, presented its materials as narrative articles about the individuals it contains. Since the 1990s, KCL's Department of Digital Humanities (formerly known as Center for Computing in the Humanities) has been involved in the development of structured prosopographical databases, and has had direct involvement in Prosopographies of the Byzantine World (PBE and PBW), Anglo-Saxon England (PASE), Medieval Scotland (PoMS) and now more generally northern Britain ("Breaking of Britain": BoB), and is currently in discussions about others. DDH has been involved in the development of a general "factoid-oriented" model of structure that although downplaying or eliminating narratives about people, has to a large extent served the needs of these various projects quite well.

DDH's factoid-oriented prosopographical model are currently all expressed using the entity-attribute-relationship model of the relational database. The structure formally identifies obvious items of interest: Persons and Sources, and extends to related things like Offices or Places. In our prosopographical model the *Factoid* is a central idea and represents the spot in a primary source where something is said about one or more persons. It links people to the information about them via spots in primary sources that assert that information. By creating "factoids" which assert things about what the source says about people, the factoid approach prioritises the sources, and our historians' reading of them. Our data about a person is not, then, so much a narrative that presents a summary written by the prosopographer (as it was in the articles about persons included in pre-digital prosopography) as a collection of information about what the sources say about him/her, and can represent the multiple, perhaps contradictory, voices of the different sources simultaneously: one saying she is a Saxon, but another saying that she was from Northumbria.

Bradley and Short (2005) has a more complete overview of the factoid model than what there is room for here.

Recent development, particularly the WWW, and its related technologies around the Semantic Web have promoted the possibility to both interconnecting dispersed data, and allowing it to be queried semantically. Central technologies to support this approach are ontology languages such as RDF and OWL. Modelling the work of prosopography in a framework such as OWL is in many ways similar to relational database modelling: including the idea of classes and slots which correspond quite closely to entities and attributes in RDB modelling, and in the handling of relationships between data.

So far, the DB structured approach has served our approach to prosopography well – producing digital resources that are well received by the research community, and the broader public. What, then, are the advantages of rendering our prosopographical material which is already highly structured into an ontology instead? First, ontology systems such as OWL, by being a part of the Semantic Web initiative, are designed to inter-operate between independent resources. Our recently started project called *The Breaking of Britain* – which will produce separate prosopographical databases of people in Scotland and Northern England, will be our first experience of the need for substantial linking and searching between them. Second, ontology systems provide a mechanism to engage more heavily first order logic in the search. (there is something of an introduction to this issue in Zöllner-Weber 2009, although in the field of literary scholarship). First order logic could help with genealogical data (where an asserting that A was son of B, also can be interpreted as stating that B was mother or father of A), and can facilitate the management of the complex date searching needs of materials from Medieval times. Finally, we expect that a full ontological expression of our approach to prosopography will provide a richer and more transparent formal expression of the semantics attached to our approach than we have at present. As a result, such a prosopographical ontology will embody a shared conceptualisation of the field useful to both computers and people (cf. Gruber, 2000).

One of the central approaches to modelling in the Semantic Web world is to develop a model that contains elements of other, compatible, schemas. Borrowed elements must not only match structurally, but the semantics of the classes and slots in the shared model have to match conceptually as well. At present, the existing ontology that we know about that best matches our interests is CIDOC-CRM (CIDOC 2006 and Doerr 2003), and it has been mapped to both RDFS and OWL (Schiemann 2010) – but it is aimed at the needs of the museum and archive community as a way of representing cultural heritage materials.

CIDOC-CRM has several advantages as the base for a prosopographical ontology: it is sympathetic with an historical view of the materials it represents, and identifies relevant entities: including persons, places, events and sources. Page 11 of the CIDOC-CRM specification points out that "The CRM does not propose a specific form to support reasoning about possible identity", but we are not asking the ontology for these projects to do that. The identity assertion is the work of our historian partners, and the point of the ontology is not to derive the identity of persons from the ontological-expressed data for us, but merely to express what the historians assert about their materials (including assertion of identity of individuals) in ways that support sophisticated searching.

There has been a stream of argument about the black-and-white nature of assertions made through computer ontologies, implying that this bi-polar nature is a significant flaw when applied to Humanities materials. Indeed, we expect that much of the discussion about computer ontologies – centered often on still the relatively simple problems within science and engineering (see, for example, the discussion in Gruber 1993) that have been used as examples – have put Humanists (even Digital Humanists) off. Veltman 2004 provides this kind of argument when he claims that the preservation of culture requires the dealing of changing meanings over different places and times, and that computer ontologies try to "create data structures that

assume a single world-view". (p. 7). Now, this would indeed be a significant concern and, not accidentally, fit with, say, Louch's (Louch 1969) reasoning about why narrative, with its subtlety of expression, remains for many historians the main vehicle for research output. However the story need not be as pessimistic as Veltman seems to believe, since ontology modelling need not mean that a single view of the material is an inevitable result.

Indeed, our factoid approach can show that formal structuring if designed correctly need not impose, as Veltman implies, a single perspective on the data it models, but is capable of accommodating a range of views from the different sources. In the factoid model statements about the people are not made baldly: "Alfred the Great learned his Latin from Plegmund, Asser, Grimbold and John". Instead, the statement records what a contemporary source says: "The source of the West-Saxon Version of Pope Gregory the Great's 'Pastor Care' (section 7.18) asserts that Alfred the Great learned his Latin from Plegmund ...". By introducing the source as an intermediary, the model can also accommodate contradictory statements from different sources.

CIDOC provides classes meant to facilitate the representation of this *iatus* between a fact and the reading (expressed by the source, or by us in virtue of our editorial role) of that fact. On page 8 of the CIDOC specifications, the E13-Attribute Assignment class is presented as what allows the "documentation of how the respective assignment came about, and whose opinion it was". By using this class all the "properties assigned in such an action can also be seen as directly attached to the respective item or concept, possibly as a collection of contradictory values". Examples of how to employ this mechanism to the end of constructing a reified description of reality can also be found in the context of the 'exhibition problem' discussion (Eide, 2009). An alternative solution could be instead extending CIDOC by using ad hoc classes; in the context of building ontologies for the philosophical domain, an approach of this kind has been previously documented by one of the authors (Pasin, 2009, p.7). Here a supposedly created 'Interpretation Event' class extends CIDOC and allows different users to organize networks of philosophical ideas in a subjective manner.

Despite some few early examples available, the issue of how to properly map our structured factoid model into an OWL ontology based on CIDOC-CRM remains quite a difficult and interesting one. In this paper we will report in some detail on our most recent attempts to address this problem, by reviewing the pros and cons of the existing approaches and exploring in some depth aspects of the mapping that required significant enrichment or extension of CIDOC to accomplish.

## **References**

Bradley, John and Harold Short (2005). "Texts into databases: the Evolving Field of New-style Prosopography". In *Literary and Linguistic Computing* Vol. 20 Suppl. 1:3-24.

CIDOC 2006. *The CIDOC Conceptual Reference Model*. International Council of Museums. <http://www.cidoc-crm.org/>

Doerr, Martin (2003). "The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata". In *AI Magazine* Vol. 24 No. 3. Online version available at <http://www.aiide.org/ojs/index.php/aimagazine/article/viewFile/1720/1618>

Eide, Oyvinde (2008). "The Exhibition Problem. A Real-life Example with a Suggested Solution". In *Literary and Linguistic Computing* vol. 23 (1) pp. 27-37

Gruber, Tom (1993). "Towards Principles for the Design of Ontologies Used for Knowledge Sharing". In Nicola Guarino and Roberto Poli (eds.) *Formal Ontology in Conceptual Analysis and Knowledge Representation*. Kluwer Academic publishers.

Gruber, Tom (2000). "Every Ontology is a Treaty". Interview for *Semantic Web and Information Systems SIG of the Association for Information Systems. SIGSEMIS Bulletin* vol. 1 (3).

Louch, Alfred .R. (1969). "History as Narrative". In *History and Theory* Vol. 8 No 1. pp. 54-70.

Martindale, J.R. (1992). *The Prosopography of the Later Roman Empire, 3: A.D. 527-641*. Cambridge: Cambridge University Press. 1992.

Pasin, Michele and Motta, Enrico (2009). "Ontological Requirements for Annotation and Navigation of Philosophical Resources". In *Synthese*. Online at <http://www.springerlink.com/content/20275389857wj5v3/>.

Schiemann, Bernard, Martin Oischinger and Günther Görz (2010). *Erlanger CRM/OWL*. Online at <http://erlangen-crm.org>.

Veltman, Kim H. (2004). "Towards a Semantic Web for Culture". In *Journal of Digital Information*. Vol 4 No 4. Online at <http://journals.tdl.org/jodi/article/viewArticle/113>.

Zöllner-Weber, Amélie (2009). "Ontologies and Logic Reasoning as Tools in Humanities". In *Digital Humanities Quarterly* Vol 3 No 4. Online at <http://digitalhumanities.org/dhq/vol/3/4/000068/000068.html>.