Binary Truths: Developing a Linked Data Model for Historiographical Arguments

M. H. Beals
m.h.beals@lboro.ac.uk
Loughborough University, United Kingdom

Introduction

This paper will argue that historical linked data should not be used solely as means of storing facts but also as a means of improving historiographical discourse by formalising rhetoric, clarifying premises and evidence, and allowing for a distant reading of historical interpretations.

The need for such formalizations can be seen in the following interaction between two historians. An argument is made by one historian with a selection of evidence to support it. Another historian responds with a counter argument and counter evidence. The first becomes frustrated; the second did not understand their argument and their new evidence is irrelevant. The second is equally frustrated as the first has not properly addressed their concerns and instead glosses over them to return to his or her original argument. To the outside observer the conflict appears intractable; neither side is willing to concede the other's points and modify their interpretation of the event or process. This is not a case of academic stubbornness; it is historiographical switchtracking.

Switchtracking (Stone and Heen, 2015) is the result of two similar but non-identical conversations taking place at the same time. In the above, both historians are discussing a single historical problem but have interpreted that historical problem in slightly different ways, responding to each other with own their interpretation of the question in mind. By leaving their specific goals implicit or relying upon ambiguous terminology they have allowed their arguments to be easily misconstrued or misidentified (Godden, 2013). For example, the historical problem "What caused the Salem Witch Trials?" might not only lead to different interpretations—community conflicts, economic disparity, rye-ergot poisoning, religious fanaticism—but also different incarnations of the question itself:

- "What were the causes of the Putnam accusation?" or
- "Why did the Salem Witch Trials begin with the Putnam accusation?" or
- "Why did the Putnam accusation escalate into a wider hysteria?"

The different interpretations that arise from these similar but non-identical questions can lead to historians speaking at cross-purposes and unnecessarily hinder our wider understanding of historical events. The simplest solution is to better define the premises and hypotheses of a given study. However, the challenge of providing a defined, testable hypothesis, combined with the semi-narrative writing style preferred by historians, often precludes this level of clarity. Limited by a perpetually incomplete evidence-base, fuzziness is assumed and allows if not promotes poorly aligned debates.

Linked data may provide a two-fold solution to this problem. Traditionally, clarity and reproducibility in historical research has relied upon three methodological pillars: the quotation, the citation, and the acceptance of interpretive interoperability. The first two work in tandem, providing clear links to or examples of the precise evidence used. Limitations of publishing space have previously reduced the comprehensiveness of these pillars but now the ability to provide targeted hyperlinks and sustainable datasets online allows for a much greater degree of precision. However, the use of page and line numbers, hyperlinks, DOIs, and other edition indicators are inconsistent across the history publications. Likewise, citation standards and allusion conventions differ between publications, subfields, and other communities of practice. It would be difficult, and arguably undesirable, to suggest homogeneity. Without this, however, it is impossible to prevent switchtracking and the misinterpretation as to which precise evidence is being used for which purpose.

The integration of a linked data layer, a meta-document attached to a piece of historical writing, could serve as a remedy to this problem. Bringing together existing ontologies for describing geographical, biographical, and chronological data as well as digital or digitised documents provides a straightforward means for creating strong, definite links between historiography and the data that underpins it. Because such a layer could vary in depth of detail, it could begin with the basic citation information expected of traditional footnoting, but flexibly add layers of detail that

would be infeasible in traditional journal or monograph typesetting.

Linked data is already in common usage in certain historiographical and heritage circles, usually in the publication of discrete datasets or in cataloguing digital, digitised and traditional archive collections (Meroño-Peñuela et al., 2013). However, their integration with specific piece of historiographical writing is more complex. At their most basic level, they differ little from traditional citation practises and the added value of precision may not fully compensate for the additional effort in producing this metadata layer. Instead, it is the interpretation of that evidence, and the analytical linkages made by historians, that provide the most significant opportunity for developing historical research. Within the historical and heritage community, linked data is often considered to be limited to 'fact-based' information and cannot convey or represent the analytical frameworks that the narrative provides. Indeed, as of writing, there appears to be only one complete ontology for expressing rhetorical logic, which is poorly maintained with no clear evidence of it being employed in academic debate (Dumontier, 2014). Moreover, it focused upon highly structured logical expression; historical research often relies upon highly fragmented data, requiring vary degrees of informed speculation, which, when undocumented, is the primary cause of switchtracking.

Creating an ontology that can provide definite relationships between evidence, premises, correlations, causations, formal logical deductions and speculative interpretations would allow historians to maintain the semi-narrative writing style expected of historical research but add a layer of unequivocal—if less poetic—statements that provide unambiguous statements of their argument and its components. Beyond serving as a mechanism for researchers to refine their argumentation, presenting a complex historiographical interpretation as a collection of interconnected relationships—literal and rhetorical—would allow for a computational comparison of arguments; similar conclusions could have their evidences combined whereas contradictory interpretations would have a clear set of evidences and premises from which to begin an investigation of divergent views.

This paper will therefore discuss the issues surrounding the creation of an ontology that combines well established ontologies regarding historical evidences and document provenance alongside rhetorical relationships between premises and conclusions. It will demonstrate how one might create a graphical representation of a narrative text and vice versa. The paper will be presented in both semi-narrative long form and RDF triples.

Bibliography

Dumontier, M. (2014). The Semantic Science Integrated Ontology http://semanticscience.org/resource/SIO_000261.rdf (accessed 1 November 2016).

Godden, D. M. (2013) Arguing at Cross-Purposes: Discharging the Dialectical Obligations of the Coalescent Model of Argumentation. *Argumentation*, **17**(2): 219–43 doi:10.1023/A:1024032009784.

Meroño-Peñuela, A., Schlobach, S. and Harmelen, F. van (2013). Semantic Web for the Humanities. In Cimiano, P. (ed), *Proceedings of the 10th Extended Semantic Web Conference, ESWC 2013, Montpellier, France, May 28-30, 2013, Lecture Notes in Computer Science*, vol. 7882. pp. 645–49.

Stone, D. and Heen, S. (2015). *Thanks for the Feedback: The Science and Art of Receiving Feedback Well*. New York: Viking.