

Creating Historical Identity with Data
A Digital Prosopography Perspective

John Bradley

Seiten 103–119

aus:

**Toward Undogmatic
Reading**

Narratology, Digital Humanities and Beyond

Marie Flüh, Jan Horstmann, Janina Jacke, Mareike
Schumacher (Eds.)

Hamburg University Press
Verlag der Staats- und Universitätsbibliothek Hamburg
Carl von Ossietzky

Printed with kind support of the *Digital Humanities Association for the German-speaking area* (DHD)

Impressum

BIBLIOGRAFISCHE INFORMATION DER DEUTSCHEN NATIONALBIBLIOTHEK

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <https://portal.dnb.de> abrufbar.

LIZENZ

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Das Werk steht unter der Creative-Commons-Lizenz Namensnennung 4.0 International (CC BY 4.0, <https://creativecommons.org/licenses/by/4.0/legalcode.de>). Ausgenommen von der oben genannten Lizenz sind Abbildungen und sonstiges Drittmaterial.

ONLINE-AUSGABE

Die Online-Ausgabe dieses Werkes ist eine Open-Access-Publikation und ist auf den Verlagswebseiten frei verfügbar. Die Deutsche Nationalbibliothek hat die Online-Ausgabe archiviert. Diese ist dauerhaft auf dem Archivserver der Deutschen Nationalbibliothek (<https://portal.dnb.de>) verfügbar. DOI <https://doi.org/10.15460/HUP.209>

ISBN 978-3-943423-87-7

COVERGESTALTUNG

Jan-Erik Stange

SCHRIFT

Alegreya. Copyright 2011: The Alegreya Project Authors (<https://github.com/huertatipografica/Alegreya>). This Font Software is licensed under the SIL Open Font License, Version 1.1. This license is also available with a FAQ at: <http://scripts.sil.org/OFL>

DRUCK UND BINDUNG

Books on Demand – BoD, Norderstedt

VERLAG

Hamburg University Press, Verlag der Staats- und Universitätsbibliothek Hamburg Carl von Ossietzky, Hamburg (Deutschland), 2021
<http://hup.sub.uni-hamburg.de>

Table of Contents

Preface	9
Introduction: Undogmatic Reading – from Narratology to Digital Humanities and Back	11
<i>Marie Flüh, Jan Horstmann, Janina Jacke, Mareike Schumacher</i>	
Narrative Motivierung	31
<i>Wolf Schmid</i>	
Erzählen vom Selbst als Bewusstsein, ein Versuch zur Weltbewältigung La débil mental (2014) von Ariana Harwicz	45
<i>Inke Gunia</i>	
An “Undogmatic” Reading of Lyric Poetry Defending the Narratological Approach to Poetry Analysis	63
<i>Peter Hühn</i>	
With the Hedgehog or the Fox?	73
<i>Willard McCarty</i>	
Über Metaphern und die Voraussetzungen für ihre Verwendung in der Informationstechnologie	81
<i>Manfred Thaller</i>	
Creating Historical Identity with Data A Digital Prosopography Perspective	103
<i>John Bradley</i>	
Algorithmen zwischen Strukturalismus und Postcolonial Studies Zur Kritik und Entwicklung der computationellen Literaturwissenschaft	121
<i>Evelyn Gius</i>	

Cesare Beccaria's Dei Delitti e delle pene (1764)	133
Approaching the Multilingual Textual and Paratextual Tradition from an (Undogmatic) Digital Point of View	
<i>Claudine Moulin, Christof Schöch</i>	
Lässt sich die Grenze zwischen Realismus und Früher Moderne empirisch bestimmen?	145
Ergebnisse und Fragen eines Eye-Tracking-Experiments mit zwei Brunnengedichten von C.F. Meyer und R.M. Rilke	
<i>Thomas Weitin, A. Vanessa Möschner</i>	
List of Figures	157
Contributors	159

Creating Historical Identity with Data

A Digital Prosopography Perspective

John Bradley

As someone who has worked throughout my career in the Digital Humanities (DH), I keep an eye on what is happening in Computer Science by regularly reading one of its major journals, the *Communications of the ACM*. While reading the recent April 2019 edition, I was surprised to find an article entitled “Identity by Any Other Name” (Helland, 2019). Up to then, I had not noticed much interest in questions of what constituted identity in Computer Science. In that world, related words like *identifier* might well be invoked frequently in certain kinds of digital development work, but their significance beyond a rather basic, formal one seemed to be little considered. As it turns out, Helland too is primarily interested in that part of identity that focuses on digital identifiers, and is thinking about identity in the context of what he calls “complex multi-company e-commerce”, with much of the article exploring what happens to digital identity when material is shared between what Helland calls “intertwined systems” (p. 80).

In contrast, identity is more likely to appear as an interesting topic for discussion in the humanities, albeit from quite a different perspective: debates about “identity politics” perhaps provide a good example. For those in the humanities, I suspect that Helland’s discussion about identity would seem to be too narrowly and formally defined to be clearly related to their interests. Still, perhaps we can find things that bring the two quite different concerns more closely together by focusing on a perspective that arises out of the Digital Humanities, which always has to struggle with reconciling the formal nature of computer representation with the nature of a humanities perspective. The Digital Humanities provides a kind of middle ground where at least some aspects of these two kinds of concerns can usefully meet.

I started work in what we now call the Digital Humanities before it had that name, first at the University of Toronto starting in the 1970s, and then, since 1997 at King’s College London’s Department of Digital Humanities (DDH). At King’s, I was

involved (with responsibilities ranging from a junior technician to principal investigator) in more than 20 projects that aimed to develop a digital representation of various humanities academic projects. Often the project work involved finding an effective balance between how our humanities colleagues felt an issue needed to be dealt with and what the technology made possible. Inevitably, sometimes questions about the handling of identity came up.

Furthermore, although I was never going to be the right person to dig into things like identity politics, it turned out that the great majority of our projects could be categorised as historical and, thus, in at least that particular humanities domain. Hence, the digital representation work for these 20 projects obliged us all to think more than once about issues of a less politically hot topic related to identity: identity of historical entities, particularly persons. Many of these projects were, or had significant components in, digital prosopography – which by its very nature is about the identification and thus the identity of historical persons – and were based on historical work covering quite a range of periods and cultures.¹ In this paper, then, I intend to take up the identity issue from a digital history perspective, and by this I mean to explore issues that are in part rooted in a Computer Science perspective, and in part on the humanities historical one.

Traditional Prosopography

The Merriam-Webster dictionary defines prosopography as “a study that identifies and relates a group of persons or characters within a particular historical or literary context”. Thus, for example, one of our prosopographical projects, the Prosopography of Anglo-Saxon England, identifies historical people who appeared in Anglo-Saxon sources. Lawrence Stone, in an article often cited when discussing prosopography, writes:

Prosopography is the investigation of the common background characteristics of a group of actors in history by means of a collective study of their lives. The method employed is to establish a universe to be studied, and then to ask a set of uniform questions – about birth and death, marriage and family, social origins and inherited economic position, place of residence, education, amount and source of personal wealth, occupation, religion, experience of office, and so on. The various types of information about the individuals in the universe are then juxtaposed and combined, and are examined for significant variables. (Stone, 1970, p. 46)

¹ The eight projects in which I have been involved are described in Bradley 2016 on the associated web page entitled ‘Factoid Prosopographies at CCH/DDh KCL’ at website <https://factoid-dighum.kcl.ac.uk/factoid-prosopographies-at-cchddh-kcl/> (Access: June 15, 2021).

Stone does not specifically use the word “identity” here, but it is in the establishing of the individuals that make up the “group of actors” he describes that the issue of identifying them arises naturally, and it is interesting that these “uniform questions” that he says are often asked (birth, death, marriage, social origins, etc.) were often some of the central topics of interest to our digital prosopographies too.

All of DDH’s digital prosopographies can trace their origins back to the pre-digital prosopographies developed primarily by British scholars in the mid-20th century. Indeed, the first of DDH’s collaborative prosopographies, the *Prosopography of the Byzantine Empire* (PBE), which was initially published on a CD in 2001 but is now available online (Martindale et al., 2001 and 2015), was undertaken after the three volume *Prosopography of the Later Roman Empire* (PLRE) which had been completed in 1992, but had been started decades earlier under the leadership of the well-known British Late Antiquity scholar Arnold Hugh Martin (A.H.M) Jones.

PBE, by virtue of being the first of DDH’s “digital prosopographies”, is something of a transitional project, with elements in its makeup that sit well within both an older, pre-digital, approach to prosopography inspired directly by the work of A.H.M Jones, but also in the approaches that were to come. Like all of DDH’s prosopographies (and, certainly, like many of the older pre-digital ones), it primarily consists of a collection of pieces of information about individuals that the historians have established existed in the time period the prosopography covers. In PBE, these historical individuals are identified using the same scheme that had worked in PLRE, via a name and a number (the number is used to differentiate all the individuals with the same name). Figure 1 shows a typical entry for a randomly chosen individual *Konstantinos 24*:

Each name-number combination asserts PBE’s view of the existence of a particular individual. However, obviously the name/number by itself is hardly an assertion about the existence of an individual that can be usefully used by others, or contested. Since the name/number by itself does not really provide an identity for this historical person, how does the rest of the entry manage it?

Like all the individuals in PBE, this entry starts with a table of classification assertions about the individual (note that many of these classification categories are those identified by Stone in his 1970 article, quoted above, as typical of the interests of prosopography). This provides more information that begins to define an identity for *Konstantinos 24*. However, I think it is arguable that the short article which follows – a narrative – which tells a kind of story about the individual is the thing that really establishes *Konstantinos 24*’s identity. It is based on material found in the historical sources, and represents PBE’s lead historian, John Martindale’s, summary of what the sources tell us about the individual.

Konstantinos 24		Prosopography of the Byzantine Empire
Sex	M	
Floruit	M/L VII	
Dates	653 (taq) / 681 (tpq)	
PmbZ No.	3715	
Variant Names	Constantinus	
Religion	Christian	
Locations	Constantinople (officeplace); Hagia Sophia (Constantinople) (officeplace); Constantinople	
Occupation	Grammatikos; Priest	
Titles	Ekdikos, Hagia Sophia (Constantinople) (office); Grammatikos (office)	
Textual Sources	Constantinople, Third Council of (Sixth Ecumenical Council), ed. R. Riedinger, Concilium Universale Constantinopolitanum Tertium, ACO II.2. 1 (Berlin, 1990-1992); also cited from Mansi XI passim (conciliar)	
<p><i>Konstantinos 24</i> was priest and <i>ekdikos</i> (defensor) of the Church of Hagia Sophia, he was a Latin <i>grammatikos</i>: Riedinger, p. 290, lines 25-27, p. 336, lines 16-18 (= Mansi XI 396, 421) (ἐπιμενεύοντος Κωνσταντίνου τοῦ θεοσεβεστάτου πρεσβυτέρου καὶ ἐκδικου τῆς ἐναντίου ἀγιοτάτης μεγάλης ἐκκλησίας καὶ γραμματικοῦ Ῥωμαϊκοῦ or similar). The old Latin version printed in Riedinger, p. 291, lines 22-24, p. 337, lines 16-18 (= Mansi XI 395, 422) reads: "interpretante Constantino deo amabile presbitero et defensore sanctissimae huius magna ecclesiae et Latino grammatico" (or similar).</p> <p>At the tenth session of the Third Council of Constantinople (the Sixth Ecumenical Council), on 18 March 681, he translated for the Council when texts of Latin Fathers, produced by the representatives from Rome, were compared with Latin versions of the texts apparently kept in the patriarchal library at Constantinople: Riedinger, p. 290 (= Mansi XI 396) (St Ambrose), Riedinger, p. 336 (= Mansi XI 422) (St Augustine). Styled a priest and Latin <i>grammatikos</i> (πρεσβύτερος τῆς ἐναντίου ἀγιοτάτης μεγάλης ἐκκλησίας καὶ γραμματικός Ῥωμαϊκός), he told the fourteenth session of the Council (on 5 April 681) how, at the request of the patriarch Paul II (Paulos 2; 641-653), he examined a Latin <i>codex</i> of the Fifth Ecumenical Council found in the patriarchal library and discovered that the <i>libelli</i> of Vigilius were not included in the acts of the seventh session; on orders from Paulos 2 he translated them into Latin (μετέφρασα Ῥωμαϊκῶν: Riedinger, p. 652, line 19) and had them copied by Sergios 12 and bound in by the calligrapher Theodoros 53; his account was verified by Sergios 12, who quoted Paulos 2 as calling Konstantinos 24 ὁ πτεῖς Κωνσταντίνος ὁ γραμματικὸς ὁ Ῥωμαϊκός (implying that at the time Konstantinos was a <i>grammaticus Latinus</i> and had not yet taken Holy Orders): Riedinger p. 654, lines 6-7. The whole episode is recounted in Riedinger, pp. 650-654 (= Mansi XI 593-596).</p>		
(Publishable link for this person: http://www.pbe.kcl.ac.uk/person/p4335)		

Fig. 1: Konstantinos 24 in the Prosopography of the Byzantine World

Martindale's previous prosopography, the print-based PLRE, also had an article for each identified person that is similar in style to those that appear in PBE, although PLRE does not have the preceding classification category data that we find in PBE. The fact that the article's style in PBE is similar to PLRE is perhaps not surprising; Martindale was one of the principal historians in PLRE, and it would be natural to think that because PBE was actually originally conceived of as a continuation of the work in PLRE, that the materials of interest to him, and therefore presented in the articles he wrote for individuals within PBE, would be similar.

History Making and Narrative

I would like to take a moment here to think a bit more about the role of the article in PBE and PLRE and how its production constitutes the work of "doing history". First of all, as has already been noted, the articles can be thought of as short narratives that describe what can be deduced about the individual's life. The assumption that this should be presented as a narrative is deeply rooted in historiography in general. Avril Cameron makes a connection between historical persons and narrative in an article in a monograph that celebrated the completion of PBE:

History-writing is made out of all kinds of components, but information about individual persons remains among the most important. A story without persons would not be history at all. And even a Marxist Historiography of class depends on persons to give it life. Prosopography – ‘writing about individuals’, or ‘the recording of persons’ – is one methodology which gathers and digests information about the individual persons who are attested in a particular historical period; as well as uncovering specific careers and relationships, it may also provide a tool for the broader detection of historical trends. (Cameron, 2003, p. xiii)

What is relevant here to the current paper is the observation, indeed assumption, that as a kind of history, prosopography is expressed through writing and is carried out through writing about individuals; about the creation of stories about persons. Cameron’s assumption that text is the main vehicle for presenting historical research is not uncommon. Indeed, we can spot a sense of discomfort for historians with materials other than text in the observation of the noted American historian David J. Bodenhamer that “Despite a flurry of interest in quantitative history in the 1960s and 1970s, historians as a group have remained more comfortable with manuscripts than databases” (Bodenhamer, 2008, p. 220). See also Alfred Louch’s claim in an article about historiography in which he proposes “to examine the technique of narrative as it is used by historians, in order to show that it is not merely an incidental, stylistic feature of the historian’s craft, but essential to the business of historical explanation” (Louch, 1969, p. 54).

We see in these articles views expressed about the place of narratives in history. I would like to claim that those in traditional prosopography would have assumed that the narratives that they wrote for their historical individuals were the way by which identities were given to the historical people the work identified. And, indeed, although in the entries for people in PBE (of which the entry for Konstantinos 24 is an example), people are also classified according to schemes that PBE developed so that groups of kinds of individuals could be found (to locate, say, all *Grammatikos*); nonetheless it was the article that actually gives each person a historic identity.

Personal Identity and “Life Stories”

The connection between narrative and personal identity that appears to apply to traditional prosopography also seems to operate, at least in some practitioners’ minds, in another discipline that focuses on human behaviour: psychology. There one finds an interest in what is called *Life Stories*, where, according to its adherents, it provides the mechanisms by which individuals generate their own identities through the telling of stories about themselves. One of the major proponents of the psychology of

life stories is Dan McAdams who in a major review article of the field observed that “[a]s personality psychologists began to turn their attention to people’s lives, they found notions such as *story* and *narrative* to be especially useful in conveying the coherence and the meaning of lives” (McAdams, 2001, p. 100). Later, in the same review article, while talking about his own earlier writings, he writes: “In his life story model of identity, McAdams [...] has argued that identity itself takes the form of a story, complete with setting, scenes, character, plot, and theme” (p. 101). Then he goes on to say that “Identity [...] is an integrative configuration of self-in-the-adult-world” and that it operates both synchronically, integrating a broad range of different “and probably conflicting” characteristics, and diachronically (over time) to also reflect changes and developments (p. 102). McAdams’s interest is in an individual’s personal stories about his/her own personal development.

McAdam has thus proposed that narratives or stories about persons are intertwined with their personal identity. From what we have seen concerning traditional prosopography, narrative and historic personal identity appears to be similarly intertwined as well. Indeed, Stone’s very first sentence in his article about the historical method of prosopography states that prosopography (a term he claims is used by ancient historians) and “collective biography” (the term used by modern historians) refer to the same kind of historical methodology (cf. Stone, 1970, p. 46). The word “biography” surely refers to the creation of narratives to deliver the history.

Narrative Disappears: Data-Oriented Prosopography

Given the apparent connection between narrative and history provided both by prominent historians and by the work of prosopographers before and during the development of PBE, it is striking that most of the prosopographical projects in which I have been involved subsequent to PBE are not based on the preparation and presentation of narratives about their historical individuals. Figure 2 gives us an example of material available from the *Prosopography of Anglo-Saxon England* (PASE) about an individual named *Agemund 1*:

Even at first glance, it is evident that there is no obvious narrative that describes what the PASE team knew about this person. Instead, one gets the impression of a list of separate small pieces of information. The way to interpret most of it is to see it as providing a set of assertions (called “factoids”, of which more later) that the PASE team have made about this person. In the area headed “Factoid List” one sees various categories of assertions listed: *Recorded Name* (the way the name was written in the primary source), *Offices* (what offices a source document asserted he held) and *Events*, which are further classified here into *Charter Witnessing* and *Grant* and *Gift*. PASE found these

Agemund 1 (Male)

Minister, fl. 1019-1024x1035; owner of land in Dorset
e/m xi

[Open All](#) [Order by Source](#) [Help](#) [Go back...](#)

FACTOID LIST

- Recorded Name (3)**
 - Agemund (2)**
 - S955
 - S961
 - Agumund (1)**
 - S959
- Office (3)**
 - Minister (3)**
 - S955
 - S959
 - S961
- Event (5)**
 - Charter-witnessing (4)**
 - S955 - **Cnut 3** granting land to **Agemund 1**: King **Cnut 3** to **Agemund 1** (Agmundr), his minister; grant of 16 hides (cassati) at Cheselbourne, Dorset: **S955** (1019)
 - S959 - **Cnut 3** granting land to Canterbury, Christ Church: King **Cnut 3** to Canterbury, Christ Church; grant of the port of Sandwich: **S959** (1023)
 - S961 - **Cnut 3** granting land to **Orc 1**: King **Cnut 3** to **Orc 1**, his minister; grant of 7 hides (mansae) at Portesham (alias Portisham), Dorset: **S961** (1024)
 - S981 - **Cnut 3** granting land to Canterbury, Christ Church: King **Cnut 3** to Canterbury, Christ Church; reversionary grant of land at Folkestone, Kent, after the death of **Eadsige 12**, the king's priest: **S981** (1016 x 1035)
 - Grant and Gift (5)**
 - Cnut 3**.S955 granting land to **Agemund 1**: King **Cnut 3** to **Agemund 1** (Agmundr), his minister; grant of 16 hides (cassati) at Cheselbourne, Dorset: **S955** (1019)
 - S955 - **Cnut 3** granting land to **Agemund 1**: King **Cnut 3** to **Agemund 1** (Agmundr), his minister; grant of 16 hides (cassati) at Cheselbourne, Dorset: **S955** (1019)
 - S959 - **Cnut 3** granting land to Canterbury, Christ Church: King **Cnut 3** to Canterbury, Christ Church; grant of the port of Sandwich: **S959** (1023)
 - S961 - **Cnut 3** granting land to **Orc 1**: King **Cnut 3** to **Orc 1**, his minister; grant of 7 hides (mansae) at Portesham (alias Portisham), Dorset: **S961** (1024)
 - S981 - **Cnut 3** granting land to Canterbury, Christ Church: King **Cnut 3** to Canterbury, Christ Church; reversionary grant of land at Folkestone, Kent, after the death of **Eadsige 12**, the king's priest: **S981** (1016 x 1035)

This site is maintained under a Service Level Agreement by King's Digital Lab

Fig. 2: Agemund 1 in the Prosopography of Anglo-Saxon England

particular kinds of assertions about Agemund 1 in the primary sources it read. Other entries in PASE have other kinds of assertions as well, and many have a much larger set of assertions associated with them. They are always presented in the PASE web application as a hierarchically organised list similar to the one shown in Figure 2.

Below each of the categories of assertions mentioned above, there is a list of sources that made them. As it turns out, all the sources for Agemund 1 are charter sources (for PASE several thousand sources were read, and about a third of them were legal charters), and they all have Sawyer Numbers – the scheme developed by Professor Peter Sawyer and published in *Anglo-Saxon Charters: an Annotated List and Bibliography* by the Royal Historical Society in 1968. Thus, all the sources which provided information about this individual are charters, and have Sawyer Numbers to identify them: specifically, S955, S961, S981 and S959.

The lack of a narrative to describe what is known about the person is one of the most striking things that separate PASE (and the other 7 of the prosopographies carried out jointly with historian partners at DDH from the 1990s to the present day)

from PBE and from older published prosopographies. In PASE, there is a one-line description associated with all historical persons (and you can see Agemund 1's as the second line of Figure 2) but nothing else with a sense of narrative that describes the individual overall. Clearly, PASE and these other prosopographies do not have a narrative as the central part of how an historical person is identified. Instead, the identity of their historical figures arises from the assembling of these short assertions – the *factoids*. This focus on these relatively simple, brief, assertions has resulted in a name for this kind of prosopography: *factoid prosopography*, which is considered to be a specific kind of *data-driven prosopography*. Data-driven reflects the fact that the research team building them thought of themselves as, most the time, collecting data in the form of factoids from their historical sources rather than writing articles about the individuals found in them.

What Kind of Data?

The reader might note the use of the word “data” here. This is a word that has a broad range of meanings, and I am using it in a particular way here that needs to be picked apart from other meanings.

First, we can find the word “data” used by Lawrence Stone in his 1970 prosopography article when he uses the phrase “quantity and quality of the data accumulated about the past” (Stone, 1970, p. 58), and elsewhere too. In all the examples of prosopographical undertakings that Stone considers, the primary sources are texts: historical sources as prose writings. Here, Stone, is calling these primary textual sources as “data”. Writing in 1970, he seems to be clearly using the word in the largely pre-computing sense which seems to have been the primary meaning then, as the “[t]hing known or granted, assumption or premise from which inferences may be drawn” (this definition from the *Oxford Concise Dictionary*, 6th Edition published in 1976). By “data” Stone means “text-based information”. Indeed, for almost all the “data-driven” prosopographies created by DDH and historian partners, the primary sources that have been read are texts too, exactly as they were for Stone. Thus, it is not in the *nature* of the primary sources that these prosopographies differ from what Stone had in mind. Instead, it is primarily the generated output from the research that is different, and the meaning of the word “data” as it is used in the phrase “data-driven prosopography” is not to be found in the nature of the primary sources, but in the way the work is done on them.

A second common assumption about materials described as “data” needs to be spelled out here too and separated from what is meant by “data” in this article. This is the assumption that data must be by its very nature primarily numeric, and is most

useful when it can be appropriately given to statistical processes for analysis. Turning source materials into numeric material that would be appropriate for statistical analysis which could then inform historical understanding has a long and somewhat chequered history. A well-known example where many historians believe that a data-statistical approach went wrong can be found in the study *Time on the Cross: The Economics of American Negro Slavery* by the economists Robert Fogel and Stanley Engerman, published in 1974 (Fogel & Engerman, 1974). According to William Thomas (Thomas, 2004), Fogel and Engerman took a limited number of US slave records, did some rudimentary statistical processing, and tried to generalise results to make broad historical statements about the nature of slavery in the United States before the civil war. Fogel and Engerman's conclusions seemed to work against the common understandings of many historians about slavery in the USA. As a result, Thomas claims that the study "soured many American historians on computational methods" (Thomas, 2004, p. 61). He goes on to say that "some British historians have viewed computing technology variously as the handmaiden of postmodernism, as a witless accomplice in the collapse of narrative, and as the silent killer of history's obligation to truth and objectivity".

Stone is perhaps less extreme in his damnation of statistical-based analysis for history than Thomas describes, but he still does seem to connect a data-driven prosopography to statistical analysis. As a result he concludes that prosopography through statistical data has limitations since "[f]or most social groups in most areas" this approach "cannot usefully be employed before the explosion of record materials in the sixteenth century, caused by [...] the growth of the bureaucratic, record-keeping nation state" (Stone, 1971, p. 58). Indeed, sometimes there can be a kind of historic source that is best represented as tabular containing columns of numbers that might provide the foundation for statistical analysis. In certain cases (perhaps, say, in studies of historical census data), the material might even be ideal for it. However, for most historical textual sources the challenge becomes getting this kind of data out of them: finding material that is complete and consistent enough to allow statistical approaches to be used appropriately.

It is striking then that, in contrast, all of DDH's "structured data" prosopographies arise out of periods of history where all the primary sources are overwhelmingly textual and are not record-oriented data, and these sources that survive only represent a small part – usually the elite – of their society. Thus, the material found in the sources for our prosopographies are generally not suitable for statistical manipulation and, broadly speaking, cannot be represented by principally numeric data. Thus, these project's data is not the same as the numeric kind represented as columns of data, apparently meant by both Stone and Thomas when thinking of the *Time on the Cross* project.

Data in the Relational Model

If this kind of numeric data is not what is meant by the term “data-driven prosopographies”, as I have used to describe DDH’s prosopographies, what is meant?

DDH’s prosopographies all have behind their public surfaces what is called a *relational database*, and relational databases have a particular way of organising the materials they hold so as to simultaneously (i) allow the digital data to explicitly represent some aspects of things that exist in the non-digital world in such a way that (ii) a large number of different perspectives can be taken on this material. A similar, if perhaps more open-ended approach is expressed by the technologies associated with what has been called the Semantic Web.

A relational database such as one behind any of our data-driven prosopographies has to be designed before it can be used, and this design can usefully be described as structured in terms of *entities*, *attributes* and *relationships*, these three words being used in a particular disciplined and formal way. The work to sort out what entities, attributes and relationships represent what is needed in a database is called “modelling”:

- Data is organised into entities that represent classes of objects that are to be represented. For a prosopography, then, one entity type is Person, but there are other entity types too, such as Source for the historical sources, or Office for the offices that a person could hold in that society. For PASE, there are over 30 entity types. The word “instance” is used for a particular entity. King Harold II (the Anglo-Saxon king at the Battle of Hastings) is an instance of a Person entity.
- Entities have attributes that hold information about their instances. For PASE, all Persons have a name and number, for example. For King Harold II (1022/3-66), his attribute Name is recorded as “Harold” and his distinguishing number (from other Harolds) is 3. In most databases, many entities have a good number of attributes. In PASE, another attribute attached to Person is their sex, e. g. “Harold 3’s sex is Male”.
- Entities may relate to each other, so, for instance, there is a relationship between the historical persons and the historical sources that mention them. These relationships are formally defined in the database model.

All the statements, including those expressing the relationships that one can make from such a system, can be expressed only in terms of very simple statements, like the one asserting Harold’s maleness. Thus, any kind of richness of expression that a data model of this kind has comes not so much from the structural nuance or subtlety of any one of them, but from the number of entities, and of the attributes and relationships between them that can be expressed. This may seem to the reader to be too

much of a compromise to suit the challenges of expressing things about history, and we will return later in this paper to this question briefly to suggest why it worked as well as it did for our prosopographies.

Semantics

This is, of course, just a very brief overview of how one needs to think when doing the modelling work involved in the creation of a database such as the one that stores materials for PASE. The important thing here for the argument in this paper is this: the model for a database, to be most useful, has to have a structure that corresponds to those things, and aspects of those things, that it is representing in the world. Many database designers use the word “semantics” to represent this relationship between the database and the world they are modelling. There are two different kinds, or levels, of semantics here. First, there is the semantics that describes how to define entities, attributes and relationships, and what the use of these definitions causes the digital technology to do: the semantics of database technology. These semantics apply to relational databases in general. The second level of semantics, and the one more interesting to us here, arises when a particular database has been modelled, and it describes the connection the particular entities, attributes and relationships have to those aspects of the world that are being represented.

Here, as is so often the case when a word first developed elsewhere is brought in to apply to something in computing, the word “semantics” has a related, but also somewhat different, meaning from the linguistic and philosophical world where the word is often seen. In those fields, the word *semantics* is generally tied to natural language: the languages that humans speak, or read and write (and perhaps in which they even think). As the prominent American philosopher Jerry Fodor said, in an interview in *ReVEL magazine*, right at the very beginning, about the significance of semantics in philosophy:

Semantics [...] is part of a grammar of [a] language. In particular, it's the part of a grammar that is concerned with the relations between symbols in the language and the things in the world that they refer to or are true of. [...] The intuition is that [...] semantics is about how they relate to their referents in the nonlinguistic world. (Fodor, 2007, p. 1)

Fodor's semantics is about how words and structures in language relate and connect to the “nonlinguistic world”. He frames a meaning for the word *semantics* by tying it to natural language, and thus for him the study of semantics comes out of linguistic studies. Later, he explores the idea that semantics is about a representation of the world:

Perhaps the most important thing to understand about the cognitive mind is that it is somehow able to represent the world. What makes that so important is that, all else equal, how one acts is determined by how one represents the world (rather than by how the world actually is.) (Fodor, 2007, p. 4)

He claims to abhor what he calls the “fashionable post-Modern relativism about truth, knowledge and the like” and is not willing to say that “there is ‘nothing beyond the text’” (Fodor, 2007, p. 7). To Fodor, our concepts about the world are constrained not only by their relations to one another in the mind, but also by their relationship to the world outside ourselves.

The use of the word *semantics* for the work of database modelling that was described earlier shares the idea of connection with things in the world, but, unlike Fodor, who is apparently interested in how semantics determines how one represents the world within the framework of natural language, the semantics in the database is not filtered through natural language as its expression. Thus, we have a kind of borrowing of a term that arose in one context for a slightly different one. What is meant by *semantics* in database modelling is not so much about written or spoken language. Instead, there is a pragmatic interest in applying the entity/attribute/relationship concepts in the model so that everyone involved in the design and use of the database agrees that, to the best extent possible, the structures they create mirror a representation to aspects of things in the non-digital world. There is in fact a common understanding between database developers (and in the so-called Semantic Web too) that the semantics behind a database’s structures are also concerned with the relationship between them and “the things in the world that they refer to or are true of”, to quote Fodor again.

Although language syntax is not a part of this database-oriented semantic analysis, individual words (usually nouns) in a spoken language are used, in that the naming of things such as the entities and attributes is part of the design of a database. Both entities and attributes have names, and one uses these names in the database model to ground, for the human user, the digital object in its corresponding non-digital item. It is the finding of these entities, attributes and relationships, naming them, and explicitly mapping them to things with the same names in the world that makes database modelling work in ways that are useful. In some sense, however, (just as in natural language) syntax provides a framework in which words relate to each other and express semantic meaning; the names of components in a database design relate to each other through the restricted formalisms of database semantics of entity/attribute/relationship and are, perhaps surprisingly, able thereby to also express useful things about the world.

Modelling History and Historical Understanding

The reason why this kind of database modelling served the interests of our historians is that the entities that were detected in the sources were not only those that are obviously a part of the historical world such as People, Sources, etc. but included entities that connected their *historical understanding* of this historical world. This is where the idea of Factoids and Factoid Prosopography arose (cf. Bradley, 2016, web page “What is Factoid Prosopography all about?”). As this website says, “[a] factoid is a spot in a source that says something about a person or persons”. In the factoid prosopography approach, an historical source is treated as a source for assertions, made by the source, about persons. The ironic character of the name “factoid” is intentional: each factoid records that a source makes a statement (not necessarily now thought of as “true”) about a person. While creating factoids, then, the historian’s interpretation is always focused on the individual source he or she is reading: “what is this source saying about this person or these persons in this spot in the source”. It is through the formalism of factoids that the connection between historical sources as text and historical people became modelled as data. This article will turn to how factoids might contribute to the issue of historical personal identity shortly. First, however, the important point here is that the factoid, as one of the semantic entities provided in the factoid prosopography model, represents the part of “the world of the historian” that sits somewhere between the historical world and the historians interpretation of it. Furthermore, the name supports the idea that the “world” being represented here is not simply an “objective” historical world of, say, the Roman Republic or Medieval Scotland. Instead, it is the more complex world of the historian, with their interpretations of what the sources mean blended with this historic, now inaccessible, world. It also, at least in part, addresses the concern that the apparently simple formalisms of structured data enforce a positivism-oriented view on the data. Factoids are not facts; they are creations of historians as a result of their interpretation of their sources.

Challenging “Metaphorical Narrativism”

Historical personal identity that emerges from a collection of factoids that are associated with an historical person, as we have seen in PASE’s Agemund 1, provides an interesting response to Alfred Louch’s claim, mentioned earlier, that “the technique of narrative” is “essential to the business of historical explanation” (Louch, 1969, p. 54). Louch is writing at roughly the time of the work of prominent history theorist Hayden White (and subsequently, Frank Ankersmit). Years later, in 1998, the history theorist Chris Lorenz wrote a critique of some of White’s observations (cf. Lorenz,

1998) focusing on what he calls their “metaphorical narrativism” model, where “the narrative form of history implies that the notion of truth as correspondence cannot be applied to historical stories in contrast with the individual statements that collectively make up these stories” (Lorenz, 1998, p. 310). Shortly thereafter, Lorenz claims that for White “the *plot-structure* of narratives” means that historical narratives should be interpreted as an “extended metaphor”: it is the historian who “*imposes* a linguistic, literary structure on the past” – even though in that past “nothing real corresponds to it” (Lorenz, 1998, p. 331). In the associated footnote he notes the difference between lives and stories: “Lives are lived and stories are told”. Thus, Lorenz says, “stories about lives must be structurally different from the lives themselves and can’t be realistically ‘copied’ in the story form”.

The advantage of the factoid approach, then, might be that it retains a set of low-level, close-to-the-source, historical assertions (even then, calling them ironically “factoids”, and recognising that they are interpretations), while avoiding the narrative-oriented construction of a higher-level “story” of the person’s life, which requires a higher level of interpretation, and is thereby subject to a more subjective rather than objective expression. In the factoid prosopographies, the historians were deriving historical assertions from the texts they found in their historical sources, and they were associating these assertion-factoids with their sense of named, historical persons, but they were not then turning these collections of factoids into stories or narratives that attempted to turn these collections into historical lives. They thus largely avoid both of the problems that Lorenz finds with White’s interpretation of conventional historical writing, which is that, in the writing of these historical narratives,

[n]either at the level of individual descriptive statements nor at the level of the narrative organization of those statements is it possible to disentangle the referential, descriptive from the metaphorical, point-of-view function, because *all* linguistic representations of reality at the same time constitute points of view at reality, recognized as such or not. (Lorenz, 1998, p. 325)

Lorenz does contest the idea that even individual statements about historical persons – here the factoids – are guaranteed, as it were, to be true: “... the idea that the truth of individual statements [...] is self-evident and beyond debate [...] cannot be upheld. At both levels the establishment of truth and falsity is dependent on fallible, intersubjective conventions” (Lorenz, 1998, p. 324). However, the factoid assertions, although created by the modern day historians of our prosopographical projects and therefore a modern interpretation of the text in the sources, seems to be accepted by many historians who have commented on the factoid model as, at least potentially,

close to “facts” as one can expect from historical sources. Lorenz’s concerns are at least apparently reduced to a kind of minimum by this approach.

As I wrote elsewhere “historical persons survive for us through their appearance in sources, and historians identify them not only by their name, but also by what they did and by other ways that they are described” (Bradley, 2014, slide 6). This statement is relevant to traditional narrative-oriented prosopography of course. Nonetheless, it seems to me that the list of factoids in displays like PASE’s Agemund 1 also generates a sense of historical identity for this individual, and that hence the work of assembling the factoids into historical, named, individuals constitutes assertions about that person’s identity. With factoids the historians are asserting something about who these people were, even though the nature of these assertions is somewhat different from what would appear if the people had been presented through articles.

Furthermore, a collection of factoids form a claim by the historians for the identity of each historical person, and such a collection is still contestable by other historians – just as people defined in narrative articles must be. It is through this contestability that a sense of “truth” for the assertion of identity arises. Lorenz notes:

The complexity of the notion of truth in the case of narratives [...] cannot be used as an argument against it, for as long as we presuppose that historical narratives refer to a real past and thus represent knowledge of the past, historical narratives constitute truth-claims that must be elucidated and not annihilated by philosophy of history. (Lorenz, 1998, p. 326)

Although arguably closer to the historical sources from which they are derived than an article about a person would be, the collection-of-factoids data approach establishes a connection with Lorenz’s sense of truth: the factoids are not meant to be read as facts, but when assembled into people, they still represent a “truth-claim” about the existence and nature of these historical individuals. Even though this expression is somewhat different in nature from the traditional history as narrative, these data-oriented prosopographies still represent an attempt to link the historical sources to a “real past” even though this can only be viewed through the glass dimly.

It is, of course, impossible to establish objectively how well our factoid-based prosopographies fit with good historical practice. Even so, it seems that our historian partners have, in general, been satisfied with their expressiveness. From time to time we hear from someone who tells us that they have been valuable to their own research. Thus, there is reason to think that this non-narrative approach to prosopography works well. They challenge the position of theorists such as Louch, Stone, White and Ankersmit that of necessity all historical work must be presented in narrative form. The approach to data we used to support these prosopographies, based

on the well-established entity/attribute/relationship approach of the relational database (and in the Semantic Web and Linked Open Data context too), has demonstrated that it has the semantic potential to represent usefully some aspects of the world, even though it is not founded on a linguistic-driven sense of semantics described by Fodor. Furthermore, when a model developed (like the factoid approach) includes not only an attempt to represent the historical world, but also some aspects of the world of scholarship about that world, the approach can create a useful representation that can reflect historical, humanities-oriented scholarship.

Obviously, this approach cannot be extended to a claim that *all* historical research can be represented without the need for narrative. Indeed, a claim might be well made that this approach only suits a small amount of what constitutes historical scholarly work. Even so, the data modelling approach when applied to an historical research agenda, suggests that there might well be, at least sometimes, a place for recognising that there are mental structures that are not narrative in nature and sit between the historian's view of their historical period of study and the prose text that they might eventually create to describe them.

Contact

John Bradley
 Honorary Senior Research Fellow
 King's College London
 Department of Digital Humanities
 Chesham Building, Strand
 London UK, WC2R 2LS
 john.bradley@kcl.ac.uk

References

- Bodenhamer, David J.: History and GIS: Implications for the Discipline. In: *Placing History: How Maps, Spatial Data, and the GIS Are Changing Historical Scholarship*. Ed. by Anne Kelly Knowles. Redlands, CA 2008, pp. 220–231.
- Bradley, John: "Ontologies for Prosopography: What's in and what's out?" Presentation for Standards for Networking Ancient Prosopographies: Data and Relations in Greco-Roman Names (SNAP: DRGN) Workshop. 31st march to 1st april 2014; <http://snapdrgn.net/wp-content/uploads/2014/04/Bradley-31-mar-2014.pdf> (2014). (Access: June 15, 2021).
- Bradley, John: Factoids: A site that introduces Factoid Prosopography; <https://factoid-dighum.kcl.ac.uk/> (2016) (Access: June 15, 2021).
- Bradley, John, Rio, Alice, Hammond, Matthew and Broun, Dauvit: Exploring a model for the semantics of medieval legal charters. In: *International Journal of Humanities and Arts Computing*, 13/1–2 (2019), pp. 136–154. DOI: <https://doi.org/10.3366/ijhac.2017.0184>.

- Cameron, Avril (ed): *Fifty Years of Prosopography. The Later Roman Empire, Byzantium and Beyond*. Oxford (2003).
- Fodor, Jerry: *Semantics: an interview with Jerry Fodor*. In: *ReVEL* 5/8 (2007).
- Fogel, Robert and Engerman, Stanley L.: *Time on the Cross. Economics of American Negro Slavery*. New York 1974.
- Helland, Pat: *Identity by Any Other Name*. In: *Communications of the ACM* 62/4 (2019), pp. 80–87.
- Lorenz, Chris: *Can Histories be True? Narrativism, Positivism and the “Metaphorical turn”*. In: *History and Theory* 27/3 (1998), pp. 309–329.
- Louch, Alfred R.: *History as Narrative*. In: *History and Theory* 8/1 (1969), pp. 54–70.
- Martindale, John: *Prosopography of the Byzantine Empire (641–867)*. First published as a CD through Ashgate International (2001), now available through the online edition at <http://www.pbe.kcl.ac.uk/> (2015) (Access: June 15, 2021).
- McAdams, Dan P.: *The Psychology of Life Stories*. In: *Review of General Psychology* 5/2 (2001), pp. 100–122.
- Stone, Lawrence: *Prosopography*. In: *Daedalus* 100/1 (1971), pp. 46–79.
- Thomas, William G. III: *Computing and the Historical Imagination*. In: *A Companion to Digital Humanities*. Ed. by Susan Schreibman, Ray Siemens and John Unsworth. Oxford 2004, pp. 56–68.