
From Index to Interoperability: The Desideratum of Authority Files in Large-Scale Digital Projects

Scholarly and Research
Communication

VOLUME 5 / ISSUE 4 / 2014

Brent Nelson

University of Saskatchewan

Abstract

Networked knowledge has long been an elusive desideratum of the digital humanities. This article argues the desirability and feasibility of linking person-entity references between a well defined and closely related set of digital projects related to early modern knowledge networks.

Brent Nelson is Associate Professor at the University of Saskatchewan. Email: brent.nelson@usask.ca

Keywords

Open data; Linked data; Indexing; Network analysis

The Web 2.0 has created a revolution in connectivity and information exchange. Instead of having to seek out information, users receive content through an ever-evolving network of interconnected people. Former classmates are reconnecting and exchanging videos of funny Christmas pajamas and puppies trying to climb stairs. A rich stream of content flows directly into the user's Facebook newsfeed, and the advertisement sidebar directs his attention to items tailored to his interests: a new diet to burn belly fat; yet another dating service. When a patron visits Amazon, she finds herself algorithmically assessed and connected with people like her who bought products she might be interested in. It has never been easier to make relevant personal connections ... unless you are Francis Bacon, Robert Hooke, Sir Thomas Browne, or John Woodward. When tracking seventeenth-century intellectual networks, the "friend of a friend" or "also interested in" algorithms won't do. This online social network won't build itself. If it is going to be built, interested scholars will have to do it themselves.

CCSP Press

Scholarly and Research Communication

Volume 3, Issue 2, Article ID 0401192, 13 pages

Journal URL: www.src-online.ca

Received August 4, 2014, Accepted September 4, 2014, Published December 18, 2014

Nelson, Brent. (2013). From Index to Interoperability: The Desideratum of Authority Files in Large-Scale Digital Projects. *Scholarly and Research Communication*, 3(2): 0401192, 13 pp.

© 2014 Brent Nelson. This Open Access article is distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc-nd/2.5/ca>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Networked knowledge has long been an elusive desideratum of the digital humanities. In the context of scholarly editing, Neil Fraistat (2012) describes a “key limitation of print editions: they are, in effect, data silos, possessing at best limited means for interoperating with a larger world of related data and tools. Similarly, electronic editions can also be data silos, and critical editions have over the past two hundred years developed highly sophisticated methods and protocols for exploiting the addressable affordances of print” (p. 331). Fraistat lists interoperability among his eight desiderata for the scholarly edition. Something approaching interoperability was one of the founding objectives of the Text Encoding Initiative (TEI). Established in 1987 by a group of forward-looking scholars, librarians, and archivists, the TEI was an attempt to create an open standard that would enable easy discourse between digital texts, but this original vision remains unfulfilled. Syd Bauman (2011) points out that, in fact, the titular term of reference in the TEI guidelines is not “interoperability,” but “interchange”: *Guidelines for Electronic Text Encoding and Interchange*. He argues that in the recent version of the *Guidelines* (P5), the term “interchange” has given way to “interoperability,” and that there is an important distinction to be observed between them. In Bauman’s definition, “an interoperable text is one that does not require any direct human intervention in order to prepare it to be used by a computer process other than the one(s) for which it was created” (n.p.). Interchange, on the other hand, more modestly entails a sharing of resources involving some form of negotiation between them. The complicating factor for interoperability, Bauman argues, is the “freedom of expression” allowed in the way the *Guidelines* are applied in a given situation; and possible range of situations (the kinds of documents, their intended audiences and uses, etc.) and possible interpretations add further complication: “Ask a room full of scholarly editors to examine a document in their field, and you will have as many interpretations (i.e., encodings) of the document as there are editors” (n.p.).

In recent years, Linked Open Data has emerged as the mechanism for the kind of interchange Bauman describes (Berners-Lee, 2006). The Alliance of Digital Humanities Organizations (ADHO), for example, recently formed a new Special Interest Group dedicated to pursuing the possibilities of Linked Open Data. This article will contextualize linked data as a mechanism for interchange by looking briefly at the development of one form of indexing in pre-digital publication. It will use this example as a basis for proposing some principles for a generalized approach to interchangeability. In brief, this article proposes to explore the possibility and potential for an authority mechanism for linking data on historical persons across diverse digital projects pertaining to the early modern period.

This article addresses the possibilities of linked data in the context of the ReKN (the Renaissance Knowledge Network), a node associated with the Advanced Research Consortium (ARC). ARC is an aggregating service, begun with NINES (Networked Infrastructure for Nineteenth-Century Electronic Scholarship) and replicated for the eighteenth century (18thConnect) and the medieval period (MESA). With respect to linking data, thus far ARC nodes have not focused on ingesting content, but rather on indexing digital objects through metadata using RDF (Resource Description Framework) and full-text searches. The ARC’s primary function is to aggregate independent projects (ingesting metadata about them so they can be searched collectively) and, increasingly, to provide search access to proprietary datasets. In ARC,

resources are connected in the form of results returned from indexing and searching, and these results come in the form of metadata records that point to the referent object. From its first incarnation as REKn (Renaissance English Knowledgebase) some twenty years ago, ReKN developed independently from ARC and has adopted a much more complex model. To simplify considerably, ReKN has sought not only to build a complex database of the kind of content ARC indexes, but also to provide links between related content, both hand-encoded and algorithmically derived (Siemens, 2010). The ambitious goal of ReKN is to provide a professional reading environment that will deliver to the user all relevant and related (digital) material connected to a primary work of interest. What is presented in this article is perhaps closer to the indexing (as in, pointing) model that is at the centre of ARC. Rather than talking about the full range of possibilities for indexing in an aggregating environment, the focus of this article is on one particular entity that is of particular relevance to a number of large digital projects dedicated to the early modern period.

My focus here is on the person entity because the possibilities of linked data are much more striking for people than, say, places – although linking places (at least geo-political places) would probably be simpler. At the cusp of modern biography, the early modern period is a sweet spot for prosopography, the study of large data sets about people and their relationships within a well-defined group or network. In contrast to the middle ages, the biographical content of the early modern period is rich enough to yield a substantial and meaningful network of relations, and yet not so vast and extensive as in the modern era. Indeed, the early modern period is when networks outside of the local and communal, on the one hand, and the political, on the other, really begin to form. One finds, for example, the network of empiric practitioners of Elizabethan London described by Deborah Harkness (2007) and the coterie culture of manuscript circulation illustrated by Arthur Marotti (1986), as well patronage circles and literary networks (O’Callaghan, 2007; Summers and Pebworth, 2000). Then there are the humanist networks of the sixteenth century and, later, the republic of letters. A major collaborative research program at McGill, the Making Publics Project (2005–2010), historicized and theorized group formations of this sort in the early modern period.¹ One particularly rich formation was the intellectual network that developed around the Royal Society of London in the seventeenth century. This is where the social network really begins to get interesting, and indeed, there are at present a number of large digital projects related to the social networks that intersected with the scientific interests of the Royal Society.

It makes sense to focus on this seventeenth-century context for thinking about the possibilities of linked data because extensive social networks can be readily identified. It also makes sense because there is already a large body of person-related data being collected, and though much of it continues to be siloed, some of it is now being unlocked. The Hartlib Papers Project (Sheffield University), for example, was recently migrated from CD-ROM and made publicly available online. There are, however, relevant projects that are far from open to linking. Here, as in other domains, content remains locked down in static formats. The Galileo Project, for example, contains a beautifully structured biography of John Woodward (a highly-connected seventeenth-century collector and virtuoso), but the project’s only format is the simple HTML file.

A number of other projects, however, are structured in an open manner that can facilitate data linking. The Mapping Early Modern London project (University of Victoria) and the Grub Street Project (University of Saskatchewan) are building databases of people mapped onto geographical locations, with data that are potentially addressable to other projects. Sir Hans Sloane's Correspondence Online (also at the University of Saskatchewan) focuses on Sloane's medical interests and his involvement with practitioners and patients. Some projects are already investigating the possibilities of linking their significant biographical data on historical persons. One of these is Early Modern Letters Online (EMLO), the flagship project of the Cultures of Knowledge Project (CKP). EMLO links biographical entities (historical people) to the Oxford Dictionary of National Biography and shares linked data with at least one other project, the Hartlib Papers Project (<http://emlo.bodleian.ox.ac.uk/about>). EMLO is also sharing metadata with the Mapping the Republic of Letters project (Stanford) to begin exploring the possibilities of network analysis applied to their letters' metadata. EMLO is adding substantially to early modern biography through its collection of metadata on early modern letters, with over 13,000 historical figures currently in their database. Perhaps even more significant than the biographical data themselves is how the data gleaned from written correspondence can track a network of relations, from the basic sender-recipient relationship, to other individuals mentioned in letters, and, in some cases, agents in the exchange process, such as couriers. With a user-contribution mechanism in place, the EMLO is positioned to be a central site for collecting prosopographical information on the literate world of the early modern period. What if these resources could be meaningfully connected to other person-rich resources for the same period?

The EMLO project provides a glimpse of the possibilities in linking references to individuals occurring in different resources. In the early modern period, we have for the first time a significant body of material that left traces of a social network in the form of letters. It is not that letters were not an important means of social engagement prior to this, but in the early modern period, letters were exchanged and have survived in sufficient numbers that we are able to track social networks on a large scale. (In EMLO there are, for example, over 6,000 letters involving physician and collector Sir Hans Sloane). These traces become more significant when the people involved with them shared a common interest or project, especially when these interests or projects involved common activities. For example, in a forthcoming article in *English Literary History*, Ruth Ahnert and Sebastian Ahnert apply quantitative network analysis to a corpus of surviving letters now held at the British Library (London) and Emmanuel College Library (Cambridge), in order to reconstruct the social network of the underground community associated with the famous martyrologist John Foxe during the reign of Mary I. This analysis allows them to measure how well-connected an individual was not simply by the number of his connections, but also by the nature of his connections: to whom he was connected, and in what role.

Another promising context to explore is that of the knowledge network around the Royal Society. A central activity of the Society was the collection of natural specimens and artifacts, the focus of my own digital project, the Digital Ark, which is a database of early modern collectors and collections of curiosities in seventeenth-century England (<http://digitalarkproject.blogspot.ca>). Fellows of the Society and their associates collected,

exchanged, discussed, and sometimes donated objects to their corporate “repository,” which was essentially a collaborative collection of curiosities. Seventeenth-century collections of curiosities were social places for people of like interests. Travellers from abroad visited known collections. Collectors collected visitors like they collected objects, often publishing names of their more socially notable patrons in their museum catalogues. The collection of objects was also a very social activity. Collectors exchanged objects, and members of the general public often contributed objects to the local virtuoso known to have an interest in strange objects. In this context, we have not only relationships of association (“friend of a friend”), but also relationships of agency and interactivity (“person X gave object Y to person Z”). This network of activity can be traced not only in correspondence but also, for example, in the *Philosophical Transactions* of the Royal Society, which included not only accounts of the meetings of the society (including, of course, the people involved) but also dispatches from people from the furthest reaches of England and the world. Another potential source of prosopographical data are catalogues and inventories, some in manuscript and others published in print, which identify the circumstances of an object’s collections and circulation.

An entry in the printed catalogue of Ralph Thoresby (1658-1725) aptly illustrates this network of collectors. Thoresby, a Leeds cloth merchant and fellow of the Royal Society, possessed a huge collection, begun by his father and much expanded through the late seventeenth and early eighteenth centuries. Much of the collection grew from donations and objects obtained through a large social network of other collectors, suppliers, generally interested associates, and others from the local region. One of the entries in his *Musaeum Thoresbyanum* (1713) reads:

The *Jaws* of a young *Shark*. Those of another somewhat larger; and the *Jaws* of a *Great Shark* (Don. Jo. *Bearcliffe* Pharm.) these are near two Yards wide; there are four, and in one Place five Rows of Teeth visible; they are white, broad and indented. A dark-coloured serrated Tooth of a Shark from *Maryland*. Don. *Ric. Richardson* M.D. A most remarkable one petrified; the bony Part is two Inches and a half long, smooth and shining, besides the Root which is rugged. Now comparing this, with those in the Head of an entire Shark, amongst the Curiosities of the Royal Society, (which are not half an Inch in the Animal that is two Yards long (d), and it will appear, that the Shark, to which this belonged, hath been above thirty Foot in Length. (Thoresby, 1713, p. 437, sig. 5T).

This catalogue entry is typical of Thoresby’s representation of his objects: the circumstances surrounding their acquisition were often as interesting to him as the objects themselves. The story of these objects is therefore incomplete without an understanding of the social network and mechanisms of exchange that supported Thoresby’s collection of them: chiefly, his letters. These are crucial to the task of fleshing out the social network of the legion of contributors to this museum. In the case of these shark jaws, John Brearcliff (c. 1618–1682), an apothecary and antiquary from nearby Halifax, presented his specimen with an accompanying letter sent to Thoresby dated August 3, 1703.² Brearcliff (spelled *Bearcliffe* in Thoresby’s entry) also turns up in a search of Early Modern Letters Online. In a letter dated May 17, 1679, he offered some Roman coins for sale to naturalist and collector Martin Lister (1639–1712).³ Lister, for his part, was Thoresby’s correspondent and a donor of objects to his museum.

Lister was himself a very active correspondent: EMLO contains over 2,600 letters associated with him, many of them involving other collectors.

Another of Thoresby's specimens was provided by physician Richard Richardson (1663-1741), also a Yorkshire man. Taking at random one of the 881 letters in EMLO involving Richardson – a letter dated February 4, 1671, to Edward Lhwyd, later keeper of the Ashmolean Museum – we find Richardson discussing the transfer of some of Dr. John Woodward's collection of fossil shells to Thoresby's museum.⁴ These fossils are recorded in Thoresby's (1713) catalogue in a section of "Fossile Shells *and* Stones of the Turbinated *Kind*" (p. 458, sig. 6Av). Thoresby (1713) records several other donations by Woodward, including a number of "lithophyta" (plants fossilized in stone), as "[g]ifts of my honoured Friends Dr. *Woodward* and Dr. *Richardson*" (p. 456, sig. 5Z2v). This is just the beginning of a very long thread of relations that can be traced from just one object in Thoresby's collection, involving a number of people known to each other in an extensive network of collectors. Many of these same people turn up repeatedly in the other databases identified above.

The current state and prospects of linked data

The case described above highlights the opportunity for, and desirability of, linking data across projects that intersect with historical networks such as those that clustered around the Royal Society. Yet there are very few, if any, examples of such linking being implemented among the many digital data sets involving the early modern period. An example of a comparatively simple set of data – coins – illustrates the difficulty of linking data. The core of numismatic (coin-related) databases has relatively few fields: issuer, mint, denomination, material, date, weight, size, obverse and reverse description, obverse and reverse inscription. In the case of a vast and well-defined domain – ancient Greek and Roman coins – the values for these first four fields are finite and easily prescribed, and the values for the next three are easily standardized. And yet interoperability between numismatic data sets remains elusive, as "differing national traditions have yet to integrate their substantial datasets on the basis of shared vocabularies, syntax and structure" (Wigg-Wolf, Meadows, & Tolle, 2014, p. 41). A roundtable addressed this challenge at the Computer Applications and Quantitative Methods in Archaeology conference in Paris in April 2014. The goal of the roundtable was to create a framework and portal that would link all digital numismatic data sets to enable "data exchange and facilitate access to data across a range of repositories" (p.41). The mechanisms have been theorized and developed using the established standard of Linked Open Data: RDF using uniform resource identifiers (URIs) to identify relationships and shared entities between data sets. Implementation has so far been less ambitious. A first attempt is being undertaken by the European Coin Find Network (CF) portal, under development at Goethe Universität, which will link three numismatic databases in Frankfurt, Utrecht, and Vienna. The CF's ultimate goal, it seems, is to provide a single site for accessing digital libraries of coins, but the degree to which these collections will be linked remains to be seen. Because the metadata for coins is finite and almost universally consistent, the prospects for adopting a common standard, one expects, are good.⁵ Nonetheless, the call for participants for the roundtable was premised on the admission that "linking disparate numismatics repositories presents a number of problems."⁶

And what has been accomplished in the TEI-XML docuverse? The answer is, it seems, not much. In response to a recent query posted by Jennifer Eustis to the TEI listserv asking for information about TEI-based projects using linked data (October 23, 2013), Eustis received few responses to the list (none of them supplying a clear example), and a few more responses off-list. If Eustis' informal poll is any indication, it seems a few projects are linking their XML documents to external reference files of their own (e.g., a custom-made database), and a few are taking steps to make their declared entities (such as people and places) "open" to the world, some going so far as to generate RDF files, though some are simply using the TEI's <listPerson>.7 It appears, then, that a few isolated cases are taking the first step of making their data open, but explicit linking to authorities remains exceedingly rare.8 I have not yet come across a successful instance of projects linked by means of authority lists of any kind.

It seems we have to look elsewhere, to an earlier time, to find a truly successful instance of linked, interchangeable data across disparate resources. In the case of the Christian Bible, a motivated community of practitioners (readers, editors, scribes, printers) developed, over a long period of time, a series of navigational aids that were useful enough to users that they were reproduced and reused. At the centre of this navigation system was a clearly defined authority for universal referencing. The essential development was a robust and universal canonical referencing system that took a central religious and cultural authority and made it into a technical authority by imposing a structure upon it that would make it universally addressable. This was a slow development, beginning with the division of the Bible into chapters by Stephen Langton, Archbishop of Canterbury, in the early thirteenth century and culminating in Robert Estienne's verse divisions in the mid-sixteenth century, which were adopted by the Geneva Bible, then the King James Version, and universally after that. By the early seventeenth century, a fully addressable Bible was a common place that new paratextual resources could point to, enabling interchangeability even without deliberate coordination between resources.

The classic case of a reading technology that takes advantage of the effective addressability of the Bible is the *concordance*. Any concordance published since the seventeenth century can be used with almost any modern version or translation of the Bible. Other Bible-referencing

Figure 1. Occurrences in the King James Bible of the word "likewise" pointing to four different Greek word sources (3668, 5615, 3898, 36).

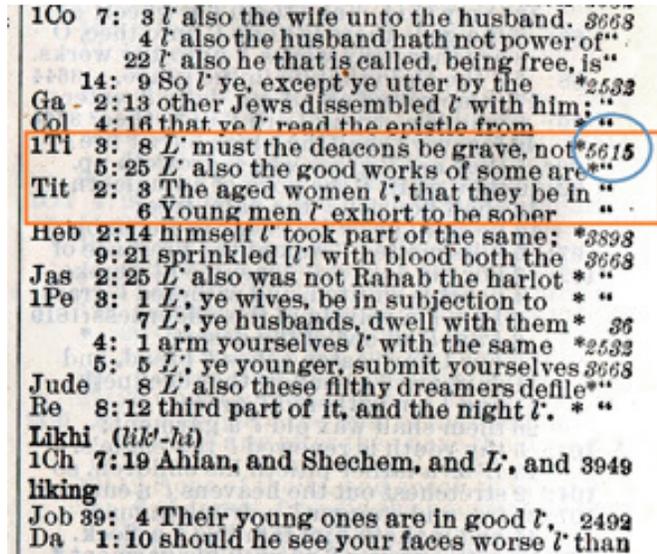
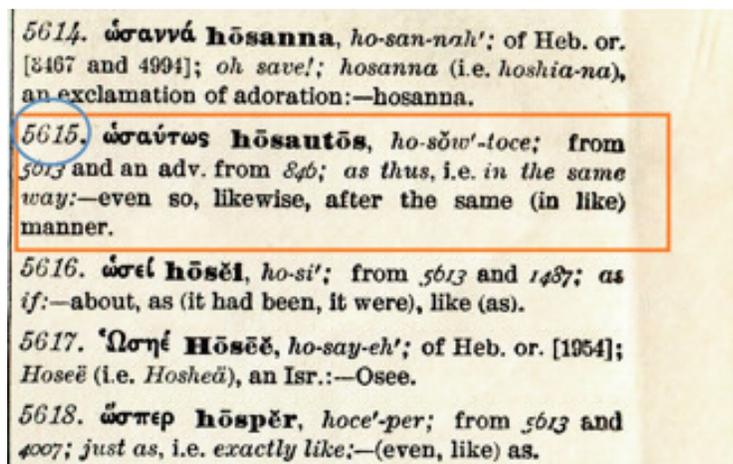
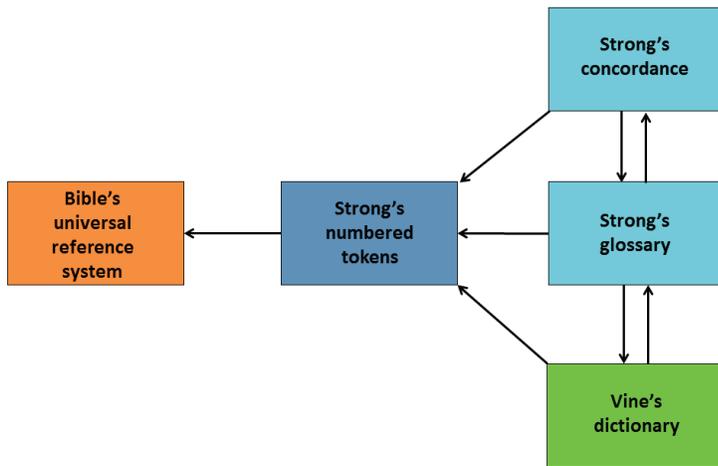


Figure 2. Strong's wordlist/glossary9



systems have been built upon this core affordance of common referencing structure. *Strong's Exhaustive Concordance to the Bible*, first published by Dr. James Strong in 1890, introduced a numeric tokenized referencing system. Every Greek, Hebrew, and Aramaic word that occurs in the Bible was given a number, so that every English word in the King James Version could be keyed by number to its original language source

Figure 3. Strong's referencing system



(by way of the universal book-chapter-verse referencing system), enabling a one-to-many relationship in cases where a single English word was used for more than one Greek word and vice versa (Figure 1). In the back of the *Concordance*, the user could locate each Greek word by number in a glossary (Figure 2). When W. E. Vine compiled his *Dictionary of the Bible*, published in 1939, he leveraged the popularity of Strong's concordance and keyed his definitions to Strong's number system, so that the reader could, with a stand-off referencing system, correlate data between the Bible, Strong's concordance, Strong's glossary, and Vine's analytical dictionary (Figure 3).

Although the case presented in this article is not easily generalizable, a few principles can be derived. The first is that an authority system works best when there is a critical mass of interested participants. The sort of co-ordination involved requires a highly motivated and co-operative community that recognizes the value and benefit of being able to link their communicated knowledge. Second, any co-ordinating system must rely on common content that can be addressed independently of any particular expression of that content. The entity (i.e., the verse) "John 3:16" could be referenced anywhere, and every user knew exactly what was being pointed to, regardless of where or how that content was being used. The chief distinction of this model is the centrality of a canonical referencing system. What makes Bible-oriented referencing and linking so effective is its addressability.¹⁰ It is not only, in the terminology of Michael Witmore (2010), a "massively addressable object" (n.p.), but a universally addressable object. Strong's ancillary system is even more generalizable in that it is not tied to a content-based structure like book-verse-chapter. While Strong's system relies ultimately on the Bible's addressability, his own number-token system for referencing functions as an independent, stand-off mechanism. The key element in his system is its indexical function, pointing to a central set of identifiers that essentially function as an authority to which other resources can point, enabling easy commerce back and forth.

The proposal

A context like the early modern knowledge network, for which there is already a critical mass of projects and content, should allow us to take some concrete first steps toward modelling linked data between projects. The key element required is a tool that would enable common referencing of authority URIs (Table 1). To begin with, libraries have already established some authority files (though, as we will see, these pose some

serious limitations). In the Anglo-American world, the principal authority is provided by the Library of Congress (LC) – even the British Library uses it. Data are presented as a Machine-Readable Cataloging (MARC) file that includes an LC control number, which is the unique identifier portion of a provided URI (Figure 4). A serious limitation of any institution-based authority is that it is incomplete and therefore inadequate for the prosopographical purposes of almost any project with extensive biographical data: libraries focus on people associated in some way with published literature. Nonetheless, it is in the library context that the first notable steps toward data linking are being taken. In Europe, one finds the CERL (Consortium of European Research Libraries) Thesaurus which, in fact, does the sort of thing proposed here. It ingests authority files from participating institutions to create a central, unionized resource, along the lines of the union catalogue. This resource illustrates the challenges of ingesting content from multiple sources. There are two CERL entries for John Woodward (the seventeenth-century collector) with slightly conflicting information (year of death), but they clearly identify the same person.¹¹ Moreover, coming from two different sources, the data presented vary, with each source providing information that the other does not. Unreconciled entities, then, pose one challenge for a collaborative authority mechanism. The most comprehensive aggregator of library-based authorities is the Virtual International Authority File (VIAF) operated by the Online Computer Library Center (OCLC), working with over two dozen participating national libraries (Canada’s not included). Again, as library-based authorities, the CERL Thesaurus, the Library of Congress Authority, and indeed all the data in VIAF, focus on entities related to publications (Figure 5). Even adding to this the National Register of Archives (the central site for British national records), a lot of biographical ground remains uncovered. Reference works with URIs, such as the Oxford Dictionary of National Biography (ODNB), could be used to supplement these lists, but some entities would still be missing. An aggregation of existing authorities is a start, but it needs supplementing.

Table 1. Data sets with entity records for collector John Woodward

VIAF	http://viaf.org/viaf/50037252/#Woodward,_John,_1665-1728
CERL #1	http://thesaurus.cerl.org/record/cnp00406311
CERL #2	http://thesaurus.cerl.org/record/cnp01341196
LC	http://ccn.loc.gov/n85195843
Wikipedia	http://en.wikipedia.org/wiki/John_Woodward_%28naturalist%29
ODNB	http://www.oxforddnb.com/view/article/29946
NA	http://www.nationalarchives.gov.uk/nra/searches/subjectView.asp?ID=P31190
EMLO	http://emlo.bodleian.ox.ac.uk/profile/person/e5e3460-b807-49a3-8d43-cff218707dd1
Digital Ark	http://drc.usask.ca/projects/ark/viewperson.php?id=31

Figure 4. Library of Congress authority file for John Woodward

The screenshot shows the Library of Congress Authority File for John Woodward. At the top, it says "The Library of Congress" and "Go to Library of Congress Online Catalog". The main heading is "LIBRARY OF CONGRESS AUTHORITIES". Below this are navigation buttons: "Help", "New Search", "Search History", "Headings List", "Start Over", "Previous", "Next", "MARC Display", and "Labelled Display". The main content area displays the following information:

LC control no.: n 85195843
LCCN permalink: <http://lccn.loc.gov/n85195843>
HEADING: Woodward, John, 1665-1728

000 00911cz a2200229n 450
001 3867827
005 20080705072115.0
008 851002n| acannaabn |a aaa c
010 __ |a n 85195843
035 __ |a (OCoLC)oca01417503
040 __ |a InU |b eng |c DLC |d DLC |d OCoLC
100 1_ |a Woodward, John, |d 1665-1728
400 1_ |a Woodward, |c Dr. |q (John), |d 1665-1728
400 1_ |a Woodward, J. |q (John), |d 1665-1728
400 1_ |a Wodward, |c Monsieur, |d 1665-1728
400 1_ |a Wodward, |c M., |d 1665-1728
400 1_ |a Woodwardus, Johannes, |d 1665-1728
670 __ |a His An essay towards a natural history of the earth ... 1723: |b t.p. (John Woodward)
670 __ |a DNB |b (Woodward, John, 1665-1728; geologist and physician)
670 __ |a InU/Wing STC files |b (usage: Dr. Woodward; J. Woodward; Monsieur Wodward; M. Wodward; ... Johannis Woodwardi ...)
670 __ |a LC data base, 9/21/85 |b (hdg.: Woodward, John, 1665-1728)
953 __ |a bz46

Figure 5. The Woodward entry from the Virtual International Authority File

The screenshot shows the VIAF entry for John Woodward. At the top, it says "VIAF Virtual International Authority File". Below this is a search bar with "Search" and "Search" buttons. The main content area displays the following information:

Woodward, John, 1665-1728
Woodward, John, médecin et géographe
Woodward, John (naturalist)
VIAF ID: 50037252 (Personal)
Permalink: <http://viaf.org/viaf/50037252>

Preferred Forms

- 100 1_ |a Woodward, John |d 1665-1728
- 100 1_ |a Woodward, John, |d 1665-1728
- 100 1_ |a Woodward, John |c (naturalist)
- 100 1_ |a Woodward, John, |d 1665-1728
- 200 1_ |a Woodward, |b John, |f 1665-1728
- 100 1_ |a Woodward, John, |d 1665-1728
- 100 1_ |a Woodward, John, |d 1665-1728
- 100 1_ |a Woodward, John, |d 1665-1728
- 100 1_ |a Woodward, John, |c médecin et géographe
- 100 1_ |a Woodward, John |d (1665-1728).

On the right side of the page, there is a network diagram showing connections between various authority files, represented by flags and icons.

Another potential authority is Wikipedia, which could potentially be a solution to the limitations described above. Wikipedia has two major advantages over the library-based authorities: there is no limiting focus, so any kind of person is included; and it is easily extensible by almost anyone. Wikipedia provides, in effect, a URI (indeed, a very intuitive and human readable one) that is able to distinguish between John Woodward the seventeenth-century naturalist and various footballers named John Woodward. In theory, any project could contribute new biographical entities to Wikipedia and thus contribute to a de facto universal authority file. The problem with this option is that a moderated Wikipedia would not tolerate entities that are not historically confirmed people. An authority for linking data from the sort of projects described above must be able to tolerate and accommodate such entities as “Captain McDougall, flourished 1680, mentioned in a letter from Edward Brown to Sir Thomas Browne.” There would, then, have to be a mechanism for declaring a “same as” relationship between entities in different projects, together with degrees of certainty. There would also have to be a mechanism for declaring a new unionized URI, once sufficient biographical data had been established and identities confirmed. The more data is aggregated and networks of relations are established, the stronger the basis for making positive identifications. Wikipedia probably will not suffice. A bespoke union authority is required.

This article has only aimed to establish the desirability and feasibility of developing a mechanism to link one type of data, one entity class, between closely related projects in a well-defined domain. In anticipation of next steps, I conclude with a top-level outline of the steps and elements required in a central authority mechanism for linking references to persons in seventeenth-century knowledge networks.

1. Ingest URIs and biographical information from CERL, LC, and other authorities, and create a new central URI for each new entity.
2. Ingest URIs from other resources with accessible data, as far as licensing permits.
3. Algorithmically connect records, resulting in a new URI.
4. Create a mechanism for associating project URIs with the new authority URI.
5. Create an import function for structured data from projects to create new entities (on the model of Zotero).
6. Create a mechanism for resolving new entities with duplicates (adapt a process pioneered by the Cultures of Knowledge Project¹²).

The challenges for such a mechanism would be administrative rather than technical. Crucially, a critical mass of data and interested and invested scholars/projects are already in place, along with a domain of knowledge well suited for this sort of index-based linking of data. The very first step, with expressed interest and support from potential stake-holders, would be a series of partnership meetings to begin working out the desired functionality and logistical requirements.

Notes

1. In one of the sub-projects of Making Publics, Matthew Milner explored ways in which such networks can be expressed computationally using RDF: <http://www.makingpublics.org/projects/?id=5>.
2. Leeds, Yorkshire Archaeological Society, Thoresby MS14.

3. Oxford University Library, Bodleian MS Lister 3 fol. 18.
4. Oxford University Library, Bodleian MS Eng. hist. c. 11 fol. 69.
5. The Numismatic Description Standard (NUDS): http://nomisma.org/nuds/numismatic_database_standard .
6. See Wigg-Wolf, Meadows, & Tolle, 2014.
7. A nice example of this approach of internal linking in XML is the Colonial Dispatches Project at the University of Victoria (<http://bcgenesis.uvic.ca>), where people mentioned in XML-encoded documents link to entities in a biography file (bcgenesis.uvic.ca/bios.xml).
8. See, as an example, Petrus Plaoul (2011-2013). *Commentarii in libris Sententiarum*.
9. Images taken from the Encourage the Young Women website, <http://encouragetheyoungwomen.wordpress.com/doing-word-studies> .
10. On the importance of addressability, see Fraistat, 2012, pp. 329-330, referencing Michael Witmore's (2010) "Text: A Massively Addressable Object."
11. See <http://thesaurus.cerl.org/cgi-bin/record.pl?rid=cnp01341196> and <http://thesaurus.cerl.org/cgi-bin/record.pl?rid=cnp00406311> .
12. As described in a conference paper by Kim McLean-Fiander at the Situating Early Modern Science Networks conference at University of Saskatchewan in April 2012.

References

- Advanced Research Consortium. *ARC*. URL: <http://idhmc.tamu.edu/arcgrant/> [August 4, 2014].
- ADHO (2014). *ADHO announces a new SIG, Linked Open Data*. URL: <http://adho.org/announcements/2014/adho-announces-new-sig-linked-open-data> .
- Bauman, Syd (2011). *Interchange vs. interoperability*. In proceedings of Balisage: The Markup Conference 2011. Montréal, 2011. URL: <http://www.balisage.net/Proceedings/vol7/html/Bauman01/BalisageVol7-Bauman01.html> . [August 4, 2014].
- Berners-Lee, Tim (2006). *Linked Data*. URL: <http://www.w3.org/DesignIssues/LinkedData.html> .
- Brennan, Michael G. (1988). *Literary patronage in the English Renaissance: The Pembroke family*. London and New York: Routledge.
- Consortium of European Research Libraries. *CERL Thesaurus*. URL: <http://thesaurus.cerl.org/cgi-bin/search.pl?start=true> [August 4, 2014].
- Cultures of Knowledge Project. *Early modern letters project*. URL: <http://emlo.bodleian.ox.ac.uk/> [August 4, 2014].
- The Digital Ark*. University of Saskatchewan. URL: <http://digitalarkproject.blogspot.ca/> [August 4, 2014].
- European Coin Find Network (ECFN)*. URL: <http://www.ecfn.fundmuenzen.eu/Home.html> [August 4, 2014].
- European Cooperation in Science and Technology (COST). (n.d.). *An interoperable supranational infrastructure for digital editions (Interedition)*. URL: http://www.cost.eu/domains_actions/isch/Actions/ISO704 [August 4, 2014].

- Fraistat, Neil. (2012). Textual addressability and the future of editing. *European Romantic Review*, 23(3), 329-333.
- Galileo Project. URL: <http://galileo.rice.edu/Catalog/NewFiles/woodward.html> [August 4, 2014].
- Grub Street Project. University of Saskatchewan. URL: <http://grubstreetproject.net> [August 4, 2014].
- Harkness, Deborah E. (2007). *The jewel house: Elizabethan London and the scientific revolution*. New Haven: Yale University Press.
- Hartlib Papers Project. Sheffield University. URL: <http://hridigital.shef.ac.uk/hartlib> [August 4, 2014].
- Hatch, Robert. The scientific revolution: A comprehensive biographical list & index of the scientific community. *The scientific revolution (1450-1750)*. URL: <http://www.clas.ufl.edu/users/ufhatch/pages/03-Sci-Rev/SCI-REV-Home/resource-ref-read/major-minor-ind/westfall-dsb/index.htm> [August 4, 2014].
- Library of Congress Authorities. URL: <http://authorities.loc.gov/> [August 4, 2014].
- Making publics 1500-1700: Media, markets & association in early modern Europe. McGill University. URL: <http://www.makingpublics.org> [August 4, 2014].
- Mapping the Republic of Letters. Stanford University. URL: <http://republicofletters.stanford.edu> [August 4, 2014].
- Marotti, Arthur F. (1986). *John Donne, coterie poet*. Madison, WI: University of Wisconsin Press.
- McLean-Fiander, Kim. (2012). Digitizing Gender: Women's Correspondence and Knowledge Networks in the Early Modern Era. Presented at *Situating Early Modern Science Networks conference*, University of Saskatchewan, April 13 and 14.
- National Register of Archives. (n.d.). *1,000 years of history*. URL: <http://www.nationalarchives.gov.uk/nra/default.asp> [August 4, 2014].
- The Newton Project. (n.d.). *Home page*. URL: <http://www.newtonproject.sussex.ac.uk/prism.php?id=1>.
- Nomismata. (n.d.). URL: <http://nomisma.org> [August 4, 2014].
- Numismatic Description Standard (NUDS). *Numismatic database standard*. URL: http://nomisma.org/nuds/numismatic_database_standard [August 4, 2014].
- O'Callaghan, Michelle. (2007). *The English wits: Literature and sociability in early modern England*. Cambridge, UK: Cambridge University Press.
- Petrus Plaoul. (2011-2013). *Commentarii in libris Sententiarum*. Critical electronic edition. URL: <http://petrusplaoul.org> [August 4, 2014].
- Siemens, Ray. (2010). Underpinnings of the social edition? A narrative, 2004-9, for the Renaissance English Knowledgebase (REKn) and Professional Reading Environment (PRE) Projects. *Online Humanities Scholarship: The Shape of Things to Come*. Connexions. Houston, TX: Rice University. URL: <http://cnx.org/content/m34335/latest/?collection=col11199/latest> [August 4, 2014].
- Sir Hans Sloane's Correspondence Online. University of Saskatchewan. URL: <https://drc.usask.ca/projects/sloaneletters/doku.php> [August 4, 2014].
- Stone, Lawrence. (1971). Prosopography. *Daedalus*, 100(1), 46-71.
- Summers, Claude J., & Pebworth, Ted-Larry (Eds.). (2000). *Literary circles and cultural communities in Renaissance England*. Columbia: University of Missouri Press.
- Thoresby, Ralph. (1713). *Musaeum Thoresbyanum*. Found in Thoresby's (1715) *Ducatus Leodiensis*, London.
- VIAF: Virtual International Authority File. (2010-2014). URL: <http://viaf.org/> [August 4, 2014].
- Wigg-Wolf, D., Meadows, A., & Tolle, K. (Chairs). (2014). *Linked data approaches to numismatic catalogues*. Roundtable at CAA2014: Computer Applications & Quantitative Methods in Archaeology conference, April 22-25, Paris. See CAA2014: Session, Roundtable, and Workshop List, p. 41. URL: http://caa2014.sciencesconf.org/conference/caa2014/pages/CAA2014_list_of_sessions_V1.pdf.
- Witmore, Michael. (2010, December 31). Text: A Massively Addressable Object. *Wine Dark Sea*. [Online essay]. URL: <http://winedarksea.org/?p=926> [August 4, 2014].