CHAPTER 4

Syriaca.org as a Test Case for Digitally Re-Sorting

David A. Michelson

1 Introduction

The decades on either side of the turn of the twenty-first century witnessed a global revolution in information technology, as exemplified by the rise of the World Wide Web, search engines, and new social media. These changes have impacted not only global communication and commerce, but have also had an effect on all areas of human culture concerned with the written word. For scholars in traditionally text-oriented fields, such as the history of Judaism or Christianity, this revolution in information technology has created a dramatic reversal in the material constraints that have long shaped their scholarship. For centuries, humanities scholarship occurred in an environment of information scarcity. In the last decade, however, scarcity has given way to a digital overabundance of information, thereby profoundly changing the material conditions for humanistic scholarship. In response to these changes, scholars have begun to construct a variety of tools, methodologies, and strategies to further scholarship in the context of a digital deluge. These competing approaches are commonly gathered under the rubric of “digital humanities.”

While it is perhaps impossible to give a precise definition to the many valences of “digital humanities,” a common ethos behind these approaches is their emphasis that humanities scholars should not merely be consumers of digital information but also active producers who are engaged in the technical, methodological, and ethical decisions that lay behind the production of knowledge in the digital age. This ethos is reflected in the three questions guiding this current volume:

---

1 A single univocal definition of “digital humanities” is not possible given the rapid growth of the field in the last decade and given the fact that humanities scholars rarely reach consensus on theoretical questions. For a sampling of recent definitions and debate, readers are referred to the opening chapters on “Defining the Digital Humanities” in Gold 2012, <http://dhdebates.gc.cuny.edu/>, last accessed May 13, 2014.
1. Do the Digital Humanities have a superficial or deep impact on research and education?
2. Are Digital Tools and Databases just accelerating research, or changing the methods and results?
3. Do the Digital Humanities offer new solutions to old problems of editing and publishing?

This essay offers some preliminary answers to these questions with reference to the specific domain of early Christian studies. My observations are drawn from the collective experience of the researchers creating *Syriaca.org: The Syriac Reference Portal*, a collaborative digital reference hub for the study of Syriac literature, culture, and history.2

The field of Syriac studies itself is at an interesting developmental crossroads. Traditionally, Syriac has received only marginal attention in the study of early Christianity. Nevertheless it offers a very rich source of materials and in recent decades has seen rapidly increasing scholarly interest and publications. This growth in the field has occurred largely in the context of the digital information revolution. Accordingly, *Syriaca.org* offers a useful test case for investigating some ways in which the digital humanities allow scholars to reimagine research methods in the history of Christianity. As a model, *Syriaca.org* demonstrates the scholarly potential of using flexible digital systems for the classification of data. In particular, two core technical frameworks implemented by *Syriaca.org*, Extensible Markup Language (XML) and Linked Open Data (LOD), offer new solutions to perennial problems of editing and publishing. Not only do these digital infrastructures dramatically increase the speed of scholarly dissemination, they have also enabled *Syriaca.org* to create new tools for collaboration and interactive scholarship. From the experience of *Syriaca.org* to-date, it is evident that the emerging tools of the digital humanities offer new paradigms for collective research and publishing in history of Christianity and other aspects of the cultural history of the ancient and medieval Mediterranean.

---

2 This article was written in 2014. Due to rapid technological change some observations may already be dated by time of publication. Nevertheless, I hope this article is of value for documenting Syriac studies’ intersection with the digital humanities. The views in this article were formulated in collaboration with many scholars including Daniel Schwartz, Jeanne-Nicole Saint-Laurent, Thomas Carlson, Tom Elliott, Aaron Butts, Nathan Gibson, Scott Johnson, and George Kiraz. The first draft of this article began as a grant application. Another version was given as a paper at the St. Ephrem Ecumenical Research Institute in 2010. I am grateful for their feedback. Any mistakes, of course remain my own.
The Digital Information Revolution

From late antiquity to the present, the plea of the New Testament author asking Timothy to “bring the cloak that I left with Carpus at Troas, also the books, and above all the parchments” has been met with a knowing sympathy from scholarly readers.\(^3\) Not only have later historians of the New Testament empathized with the author’s separation from his library, but they too would like to have access to whatever particular books and parchments the author had in mind! Access to texts has been a foundational concern of historical research, both access to primary sources and to previous research. For most of its history, European and American historical scholarship has labored to overcome obstacles to access.\(^4\) The steady increase of scholarly publishing over the past two centuries and the growth of the major European and American research libraries in the twentieth century were the first fruits of that labor. In the twenty-first century, the digitization of these same libraries and the birth of new digital publications is the full harvest, dramatically removing barriers to access.\(^5\)

Although it would be a naïve and utopian misunderstanding of the historian’s craft to assume that the historian’s desire for access to texts will ever be completely satisfied, it is worth observing that the advent of the digital library and other digital resources has profoundly altered the historian’s relationship to information. The common prior condition of information scarcity has given way to overabundance. The same digital technology which has removed many barriers to access has brought with it new challenges for the discovery of existing information. Even as the increasing availability of digital information makes access easier, that same increase makes finding relevant information or sorting information even more difficult. Moreover, this tradeoff is not complete or univalent. In some ways the rise of these new challenges of discovery

---

3 Timothy 4:13, New Revised Standard Version.
4 A general overview of information revolutions from the dawn of writing to the end of the last millenium can be found in Fang (1997). Fang’s narrative is, however, summary and subject to some criticism. A similar and more successful recent approach in brief is Liu 2007, <http://www.digitalhumanities.org/companionDLS/>, last accessed November 17, 2014.
5 I do not mean to assert, however, that this change has occurred universally or without antecedents. For many humanities scholars the majority of these new information resources remain inaccessible due to economic, social, or linguistic barriers. Moreover, while the growth in the last decade has been exponential it was not without precedent in the last century, both in the general growth of scholarly print resources in the twentieth century and also the creation of digital resources before the birth of the World Wide Web. Nevertheless, the rise of the Internet and sophisticated search engines since the 1990s has accelerated and magnified what was already an existing trend toward increasing availability of scholarly publications.
also exacerbate existing concerns about disparity in levels of access to information between scholars in high-income economies with extensive digital resources and scholars in low and middle-income economic regions.6

Scholarly response to (or even awareness of) this shift has been slow and varied. Although, nearly all humanities scholars now rely on some form of digital tools for their work (in the form of personal computers, word processing, or the Internet), a much smaller number of scholars have engaged in critical reflection on how the information revolution has changed the parameters of their inquiry.7 Nevertheless, in some disciplines or subject areas interest in digital tools and methods has been quite high. In particular, both classicists (scholars of ancient Greece and Rome) and medievalists (especially Europeanists), have been among those leading innovation and theoretical discourse in the digital humanities.8 In dialogue with scholars in these two cognate fields, this essay offers some observations on how the dynamics of the information revolution have affected a particular niche of early Christian historiography, the field of Syriac studies.

3 Growth in the Field of Syriac Studies

The Syriac language is a late-antique dialect of Aramaic that flourished as an international language of trade, culture, and religion for roughly the first millennium of the current era. During this period, Christian communities stretching from the Eastern coast of the Mediterranean to southern India and into central Asia and Mongolia used the Syriac language or were influenced by the spread of Syriac culture and literature. Today there is a growing international audience curious about the history of Syriac.

Until the late twentieth-century published resources for the study Syriac were held only by a few academic or ecclesiastical libraries. Syriac scholar

---


Kristian Heal has noted that “The history of Syriac studies could well be told in terms of the scholars’ search for, or frustrated separation from, the books they wish to study” (Heal 2012, 65–78). This observation holds true for the vast majority of the history of Syriac scholarship. In the last generation, however, the nature of this “search” has changed. Over the last quarter century, Syriac studies has benefitted from the confluence of two trends, a dramatic increase in scholarly publishing related to Syriac and the general information revolution of the Internet age. The flourishing of Syriac scholarship is perhaps most evident in the rise of new doctoral dissertations that have incorporated Syriac materials or topics. A keyword search for mention of “Syriac” in the abstracts of North American dissertations indexed in the ProQuest Dissertations & Theses Database reveals 190 Ph.D. theses related to Syriac since 1902. Of these 190, 125 were written in the last twenty-five years and 85 of those have been completed since the year 2000. If one expands the search for theses that include the word “Syriac” anywhere in the “full text” of the thesis, the results are even more pronounced with 4390 theses found of which 3593 were produced since 1990. In other words, half of all of the North American dissertations touching on Syriac topics have been written in the last fifteen years. A similar increase is found searching theses completed in the United Kingdom and Ireland and similar results may be indicated for theses in India and Europe.

Besides new research, the last quarter century has also seen a flourishing of new publications for Syriac studies. The Gorgias Encyclopedic Dictionary of the Syriac Heritage edited by Sebastian Brock, Aaron Butts, George Kiraz and Lucas Van Rompay in 2011 has provided the first comprehensive reference work for


10 North American theses were found using the ProQuest Dissertations & Theses Database Full Text, <http://search.proquest.com/pqdtft>, searches undertaken May 7, 2014. It should be noted that a “full text” search is not nearly as an accurate indicator of content as searching the abstract. It does not seem likely that all 4000 theses actually touch on matters related to the Syriac language. The author did not, however, examine all 4000 theses.

11 A search for “Syriac” in the Index to Theses (<http://www.theses.com/>) returned 79 theses from the United Kingdom and Ireland with the term “Syriac” in their descriptions, of which 25 had been written since 1990. It is difficult, however, to tell how exhaustive the Index to Theses catalogue may be. A search for theses in India in the Shodhganga repository of theses (<http://shodhganga.inflibnet.ac.in:8080/jspui/>) did not return a complete data set that could be analyzed by date, but search results did show signs that theses on Syriac topics are also flourishing in India. A search for theses in continental Europe did not return sufficient results for analysis (see the DART-Europe E-Theses Portal, <http://www.dart-europe.eu/>). Searches undertaken May 7, 2014.
the Syriac heritage (Brock et al. 2011). Other recent reference works include Michael Sokoloff’s revision of Brockelmann’s *Lexicon Syriacum* (2009) and W. Klein’s *Syrische Kirchenväter* (2004). Besides these publications a number of journals and book series related to Syriac studies began in same period, such as the St. Ephrem Ecumenical Research Institute’s journal *The Harp* (1987–present) and its monograph series *Mōrān ʾEthʾō* (1997–present).12 Similarly, the publications of Gorgias Press and Beth Mardutho: The Syriac Institute have also added to the available resources, including the journal *Hugoye* (1998–present) and several series of editions and translations such as Texts from Christian Late Antiquity (TECLA, 2006–present) and *The Antioch Bible* (2012–present).13 In addition, scholarship in Europe has also seen continued publishing of many of the long-standing resources for Syriac studies such as the *Peshitta* project at the Universiteit Leiden.

While the majority of these new publications have taken traditional print formats, there has also been a flourishing of electronic resources, perhaps most prominently the electronic format of the *Hugoye* journal (which in 1998 was a very early entrant into the world of open-access online journals).14 In addition, Beth Mardutho also sponsors an electronic mailing list also titled *Hugoye* that at present has over 600 subscribers and has had approximately 6000 posts since 1998.15 Recent publications by George Kiraz and Kristian Heal have documented a number of Syriac digital projects, again noting dramatic growth since the year 2000 (Kiraz 2007, Heal 2012).16 A partial list of resources published or in preparation include:

- *A Comprehensive Bibliography on Syriac Christianity* (The Hebrew University of Jerusalem)
- *Comprehensive Aramaic Lexicon* (Hebrew Union College)
- *eBethArké Digital Library* (Beth Mardutho and Rutgers University)
- *E-ktobe* (Centre national de la recherche scientifique, Paris)

---

12 See the history of St. Ephrem Ecumenical Research Institute’s publications in Brock 2010.
15 See “Hugoye-List: Syriac Studies Group,” <https://groups.yahoo.com/neo/groups/hugoyelist/info>, last accessed May 7, 2014. Technically the number of subscribers reflects the number of e-mail addresses subscribed which is not the same as the number of individual persons. The present author, for example, has more than one e-mail address (some of which are no longer functional) on the list.
• Peshitta Electronic Text Project (Universiteit Leiden)
• Prosopography of the Babylonian Magic Bowls (University of Southampton)
• Syri.ac (University of Oklahoma)
• Syriac Electronic Corpus (Brigham Young University)
• Syriac Studies Reference Library (Brigham Young University)
• The Harrak Collection of Iraqi Syriac and Garshuni inscriptions (University of Toronto)
• The Oliver database (Hill Museum and Manuscript Library, St. John's University)
• The Syriac Gazetteer (Syriaca.org, Vanderbilt University)
• Vatican Syriac Manuscripts (Brigham Young University)\(^17\)

In addition to these specifically Syriac-oriented electronic resources, Syriac studies has also benefitted from the general proliferation of digital information tools useful for research in the humanities such as Google Books, The Internet Archive, or HathiTrust. National and university libraries have now also made material relevant to the study of Syriac available online, including the Bibliothèque nationale de France, the British Library, the Library of Congress, the Bodleian Library, and the Universitäts- und Landesbibliothek Bonn.\(^18\)

Given this proliferation of online resources, it is perhaps not hyperbole to say that in terms of raw volume contemporary students of Syriac have access through the Internet to a greater number of books, reference works, and even copies of manuscripts than scholars have had at any previous time in the entire history of Syriac literature.\(^19\) While such a development is a beneficial one for the student, it can also be disorienting since it is a consequence of a fundamental shift in the nature of humanistic scholarship. The primary concern of the scholar is increasingly less about the scarcity of information (though much essential work remains to be done to preserve and disseminate rare Syriac

\(^{17}\) See “Appendix A” for the URLs for these resources. A similar list is published in Heal 2012, 76–78.

\(^{18}\) This list is representative, it is impractical to list all relevant digital libraries. See the discussion in Heal 2012, 68.

\(^{19}\) It is of course impossible to completely know the extent of some of the great libraries in the history of Syriac literature, such as that collected in the tenth century by Moses of Nisibis. Nevertheless, the point remains that even the most beginning student interested in Syriac has at his or her fingertips digital access to an unprecedented number of resources. Moreover, as will be noted in the section below, although the current student of Syriac has access to more resources, several core or basic resources remain woefully out of date. For an analysis see Van Rompay 2007, <http://www.bethmardutho.org/index.php/hugoye/hugoye-author-index/413.html>, last accessed May 7, 2014.
materials). Today, scholars of Syriac must also learn how to handle the rising overabundance of information. For example, a search for the keyword “Jacob of Edessa” in Google Books returns about 47,000 books. Because it is not feasible for a scholar to browse all 47,000 titles, it is clear that the increasing flow in digital information, while beneficial, brings with it new challenges for scholarship.

4 “Everything is Miscellaneous”: Old and New Challenges for Syriac Studies

It is imperative that the academic community find strategic ways to curate or channel the new flood of digital information. The field of Syriac studies has seen recent growth, but also faces significant obstacles in terms of organization of information. Some of these challenges are long-standing and well-known issues of access or publication. Other related challenges are just emerging due to the growth in the field described above and worth considering here for the first time. The inter-relationship between these “old” and “new” challenges is striking. In spite of the proliferation of publications, Syriac studies still lags a half-century or more behind related fields of research (such as Greco-Roman or Arabic literature). As Van Rompay noted in 2007, “Admittedly, an impressive number of new texts have been published in the recent decades, but when it comes to the basic tools of language and literature, it is difficult to argue that the present-day student is much better off than her or his fellow students of eighty or hundred years ago” (Van Rompay 2007, 27). Van Rompay offered a list of desiderata under the rubric of “basic tools” and first steps toward some of these items have since come to fruition (the publication of Sokoloff’s Lexicon and Van Rompay’s own work on the Gorgias Encyclopedic Dictionary of the Syriac Heritage). Nevertheless, we are still lacking many of the basic tools he

20 See Heal’s useful comparison of the modern researcher with the complaints of Hunayn ibn Ishaq in the ninth century CE (Heal 2012, 66).
22 A basic introduction to how the digital information revolution is impacting the humanities can be found in “Humanities to Digital Humanities” in Burdick et al. 2012, 1–26, <https://mitpress.mit.edu/sites/default/files/titles/content/9780262018470_Open_Access_Edition.pdf>, last accessed May 9, 2014.
highlighted, especially the extensible electronic encyclopedia he described. For example, the primary reference work on the history of Syriac literature, Anton Baumstark’s *Geschichte der syrischen Literatur*, is now more than 90-years out of date (Baumstark 1922). In spite of its growth, Syriac studies still lacks basic reference works for what Van Rompay called “core fields” within the discipline.

This “old” challenge (the lack of core reference works) is exacerbated into a “new” challenge by the growth in publishing and research. Today, Syriac studies needs up-to-date core reference works more than ever before to help scholars become familiar with the growing literature in the field. The lack of reference works holds back current scholarship in several ways. First, because much of the recent research on Syriac is only published in specialist publications (as opposed to reference works) it has remained largely inaccessible and unknown to scholars in cognate disciplines (Byzantine studies, Islamic studies, etc.). Our knowledge of Syriac language, literature, and cultures has grown but non-specialists have little access to this knowledge. Second, the lack of up-to-date reference works has led to a lack of standardized terminology making it difficult to track new discoveries. Names of authors, texts, places and even styles of Romanization of Syriac words vary widely in the literature. The lack of controlled or standard vocabulary has meant that authors, texts, and toponyms have been identified (and mis-identified) in multiple ways. Moreover, scholars and students within the field lack reference works for keeping a pace with the current state of the research on many questions in the field. As its volume increases, the literature in the field is becoming increasingly difficult to navigate. The lack of reference works then becomes a problem not only for specialists but perhaps even more so for scholars outside of Syriac studies and the general public.

---

24 Of particular note here is Van Rompay’s evocative call for an electronic expansion of the *Gorgias Encyclopedic Dictionary of the Syriac Heritage*: “we should keep working on a larger electronic database, in which many of the existing articles can be introduced as a starting point, but which should be expanded in the coming years. The more comprehensive electronic database can be more detailed and can include discussion of existing scholarship, whereas the printed dictionary will be more succinct and only provide basic information” (Van Rompay 2007, 23).

25 Van Rompay himself specifically identified lexical tools, textual corpora, and an encyclopaedia under this heading. See Van Rompay 2007, 27–28.

26 See the commentary on this problem in Riedel 2012.

27 See for example the case of the three (or more!) figures who have been blurred by tradition into the persona Isaac of Antioch: Mathews 2002.
The problem posed by the lack of basic reference resources also makes new research in Syriac difficult. The study of Syriac manuscripts provides a useful case study for these difficulties. Currently, although a scholar may find what she suspects is a previously unknown text, without a handbook of authors and works there is no standards based method to verify or disseminate the discovery. Similarly, a scholar wanting to compare multiple texts or authors has no easy way to discover what texts have been published. Moreover, for scholars wanting to compare the contents of unpublished manuscripts the difficulties are even greater as there is no union catalogue for Syriac manuscripts. In this latter case, i.e. working with Syriac manuscripts, the need for better access and discovery tools is easily illustrated by the complex workflow required to locate and compare manuscripts in two of the better documented and most important collections in academic libraries, the 1,075 manuscripts held in the British Library and the several hundred manuscripts held by the Vatican Library. Neither collection has a unified catalogue. One must use multiple published catalogues in Latin, French, and English to cover the entire collection: five catalogues for the Vatican library and three for the British Library. Moreover, a significant portion of each collection is covered only in catalogues dating from the eighteenth and nineteenth centuries that are of high quality but employ varying classification schemes.

For the patient scholar, the rewards of using these catalogues are great, but the eclectic nature of the finding aids makes searching for manuscripts containing a particular author or text laborious and the reconstruction of relationships between manuscripts tenuous. The complexity of using these collections (which have relatively strong finding aids) pales in comparison, however, to the task of locating those Syriac manuscripts preserved outside of the major academic libraries, especially those manuscripts held in ecclesiastical and private collections. If one were to attempt an exhaustive search it would require consulting several hundred small finding aids. Alain Desreumaux and Françoise Briquel-Chatonnet have published an annotated bibliography of 858 possible items. As their bibliography demonstrates, if a catalogue

---


29 Desreumaux and Briquel-Chatonnet 1991. It should be noted that not all of the items in this Répertoire are catalogues or finding aids, some are publications dealing with specific
exists for such collections (which should not be taken for granted) getting at the information is still dependent on finding access to a variety of international publications (e.g., *Journal of the Iraqi Academy Syriac Corporation*) and a broad linguistic competence (ability to read Russian, Arabic, Malayalam, etc.). In summary, the lack of accessible core tools for discovery means that these sources remain unused and unknown.

This problem of discovery of items in Syriac manuscripts is both getting easier and harder with the growth of the field and the emergence of digital projects. On the one hand, new digital cataloguing projects abound. Over the past decade the HMML library has digitized more 25,000 manuscripts from the Middle East and India (in Syriac and other languages) (Stewart 2013, 2–6). Of these, Adam McCollum and others have catalogued over 1000 Syriac manuscripts into HMML’s *Oliver database*.30 The *e-ktobe* project in France has catalogued over 1500 documents (mainly manuscripts) online and there are other projects in Lebanon, Kerala, and elsewhere.31

The proliferation of catalogues without the use of a common authority file brings with it the proliferation of different (and potentially incompatible) standards for finding and organizing knowledge in the field of Syriac studies. As we have noted already, this problem is in part a consequence of the digital revolution in information and its concomitant flood of data. In the field of sociology of knowledge, one theorist, David Weinberger, has described this problem as “the new digital disorder” where “everything is miscellaneous” or at least seems miscellaneous because the volume of information exceeds the capacity to organize it (Weinberger 2007).

5 Digital Humanities Solutions

Although the information revolution of the twenty-first century has created new challenges, few would argue that this revolution is primarily a negative
development for the field of Syriac studies.\textsuperscript{32} In the first place, the “digital disorder” also offers immediate benefits such as increased access to knowledge that was previously accessible only through a few elite research libraries. Moreover, by their very disruptive nature, the technological changes in the production of knowledge also open up new ways of organizing and advancing knowledge. From an historical perspective the shifts from scroll to codex or from manuscript to print offer analogies for how the nature of scholarship has been changed in the digital age in which scholarship is moving from page to screen.\textsuperscript{33}

In light of these changes, the field of Syriac studies can greatly benefit from joining the long running conversation within the field of the digital humanities on how scholars may best harness the powerful cultural disruption of the information revolution.\textsuperscript{34} Such engagement with the “digital humanities” should, however, begin with the caveat that it is also perhaps impossible to reduce the wide variety practices now grouped under the heading “digital humanities” down to a single definition. We may nevertheless start with Matthew Kirschenbaum’s observation: “At its core, then, digital humanities is more akin to a common methodological outlook [on the intersection of technology and the disciplines of the humanities] than an investment in any one specific set of texts or even technologies” (Kirschenbaum 2012).\textsuperscript{35} Similarly, it may be productive to conceive of the digital humanities as a shared set of questions rather than a specific set of answers.\textsuperscript{36} These questions are provoked by a common set of material circumstances (the advent of digital media) faced by scholars. These research questions not only continue traditional scholarly lines of inquiry but also to seek to re-frame research using methods that have only made become possible after the advent of digital tools and methods.


\textsuperscript{33} For an overview see Liu 2007.

\textsuperscript{34} Gold, ed. Debates in the Digital Humanities provides a number of competing definitions of the “digital humanities.” In a fitting (and entirely self-referential manner), the digital humanities project by Jason Heppler, “What Is Digital Humanities?” offers over 500 answers to its eponymous question culled from digital humanist scholars postings on the web (Heppler, <http://whatisdigitalhumanities.com/>, last accessed November 18, 2014).


\textsuperscript{36} For further definitions see the essays in “Part I: History” of R. Siemens and Schreibman 2007, <http://www.digitalhumanities.org/companion/>, last accessed November 18, 2014.
Such a focus on determining how emerging scholarly needs are mediated by digital technology (both caused and solved by digital media), offers a particularized and “local” definition of digital humanities which may prove more useful than a more global attempt at summing up the entire field. Accordingly, the discussion that follows relies upon an ad hoc definition of the digital humanities arising from the digital work on Syriaca.org and oriented specifically to the pressing needs of Syriac studies.\[37\] In this regard, our engagement with the digital humanities would perhaps best be defined as falling within the category of “digital data curation,” an area of digital work has been presented as one of increasingly urgent concern for scholars.\[38\]

In sum, while “digital humanities” as a term is relatively new, the application of digital tools to the study of humanities is not.\[39\] In fact many of the early projects in “humanities computing” (as it was first called) were in religious history fields. For example, Fr. Roberto Busa (1913–2011) and the women of his research team are widely considered to be founders of “humanities computing” through their pioneering work to encode the *Index Thomisticus* in which they created a digital index of topics in the works of Thomas Aquinas using

\[37\] In this approach I am indebted here to Claire Clivaz, and her localized analysis of digital approaches to New Testament studies in the article: “Internet Networks and Academic Research: The Example of New Testament Criticism” (Clivaz 2013). For example, Clivaz draws on Nielsen (2012), which leads her to observations about big data and digital curation which are echoed very strongly in my present essay.


\[39\] Solving the problem of information overload has long taken a central place in the critical discourse of the digital humanities. As early as 1945, scientist Vannevar Bush published a futuristic essay in which he both lamented the state of information overload (in 1945!) and imagined future forms of networked reference systems which anticipated the Internet: “There is a growing mountain of research. But there is increased evidence that we are being bogged down today as specialization extends. The investigator is staggered by the findings and conclusions of thousands of other workers—conclusions which he cannot find time to grasp, much less to remember, as they appear. Yet specialization becomes increasingly necessary for progress, and the effort to bridge between disciplines is correspondingly superficial... Wholly new forms of encyclopaedias will appear, ready made with a mesh of associative trails running through them, ready to be dropped into the memex and there amplified... The historian, with a vast chronological account of a people, parallels it with a skip trail which stops only on the salient items, and can follow at any time contemporary trails which lead him all over civilization at a particular epoch” (Bush 1945, <http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/309881/>, last accessed May 13, 2014).
IBM punch cards over a period of two decades (from the 1940–60s). George Kiraz has documented similar early moments in the application of digital tools to the study of Syriac, including how a researcher at UCLA in the 1960s had encoded Brockelmann’s *Lexicon Syriacum* on to a mainframe computer (Kiraz 2007, 37–38). What has changed since these early days of humanities computing, is the critical mass of scholars now engaged in sustained reflection on the intersection of digital information and the humanities. Melissa Terras and other scholars have documented a dramatic increase in scholars identifying themselves as digital humanists, with a sharp upward spike noticeable in 2011. According to centerNet (a network of digital humanities centers at universities and elsewhere) there are now over 125 digital humanities centers in more than 25 countries worldwide. The 2011 survey of the state of Cyberinfrastructure for the field of Classics by Alison Babeu catalogued enough different projects and resources to group them into over fifty different categories requiring over 150 pages of description (Babeu 2011, 7–174). The rise of digital humanities in prominence, and the ubiquity of technology in modern research, has lead some scholars to conclude the digital humanities is not merely a *niche* methodology or approach but is now in fact a primary mode of humanities scholarship. If it is indeed true that “We are all digital humanists now”, then it is worth asking how the field of Syriac studies can benefit from the advances of the digital humanities.

---


42 See <http://digitalhumanities.org/centernet/centers>, last accessed May 12, 2014.

43 For example: “The digital humanities are not some flashy new theory that might go out of fashion. At this point, the digital humanities are The Thing. There’s no Next about it. And it won’t be long until the digital humanities are, quite simply, ‘the humanities’” (Pannapacker 2011, <http://chronicle.com/blogs/brainstorm/pannapacker-at-mla-digital-humanities-triumphant/30915>, last accessed May 12, 2014).

6 Foundational Digital Tools: Extensible Markup Language

Without needing to take sides in a debate over whether all scholarship must be
digital, it is not controversial to assert that Syriac studies as a field can benefit
from the emerging field of digital humanities. As Kristian Heal has noted con-
cerning the emergence of digital tools, “the Syriac scholar would do well to be
informed, to make use of the emerging resources, and be involved in shaping
the tools that will best enhance the way we work” (Heal 2012, 66). Following
just such a strategy, a number of scholars have been collaborating together
since 2010 to create a suite of digital standards for the field of Syriac studies as
part of Syriaca.org: The Syriac Reference Portal.

Syriaca.org is a collaborative online hub for linking resources for the study
of Syriac. This hub is not intended per se to be an exhaustive reference work.
Instead, the primary goal of Syriaca.org is to serve as an online starting point to
facilitate the discovery of both primary data and previously published research.
Two digital technologies (or ways of structuring data) serve as the anchors for
Syriaca.org: Extensible Markup Language (XML) and Linked Open Data (LOD).
Both of these technologies are already deployed widely on the Internet for a
broad variety of purposes (well beyond those of digital humanists). Both tech-
nologies have large and sustainable user communities among humanities
scholars who are actively working toward expanding their long term utility.45

Extensible Markup Language (XML) is a data format or information stan-
dard defined and maintained by the World Wide Web Consortium (W3C),
the governing body responsible for the architecture of the Internet. The W3C
defines XML as “a simple, very flexible text format…. Originally designed to
meet the challenges of large-scale electronic publishing, XML is also play-
ing an increasingly important role in the exchange of a wide variety of data
on the Web and elsewhere.”46 XML may be familiar to many because of its
similarity to HTML (HyperText Markup Language) the data standard used
to create web pages. Like HTML, XML wraps simple text in “tags” or “ele-
ments” as a way of adding computer-processable information to text. For
example, the W3C offers this sample set of tags for how to encode a greeting:

---

45 For an overview of the current state of data modeling and knowledge organization in the

46 “Extensible Markup Language (XML),” <http://www.w3.org/XML/>, last accessed May 13,
2014.
“<greeting>Hello, world!</greeting>.” Using these tags, the user is able to indicate to the computer information about the text, e.g. that it is a greeting.

Working within the framework of XML, a number of humanities scholars have built an entire XML vocabulary for describing texts and other data sets of interest to humanists. This sub-set of XML is the Text Encoding Initiative (TEI). The TEI guidelines contain approximately 550 elements which can be used to describe texts. A wide variety of genres have been encoded using TEI from prose to poetry, from epigraphy to critical editions of manuscripts. An example of how TEI tags can be used is seen in the following simple example, a short prose description of Edessa which Syriaca.org adapted from an entry in The Gorgias Encyclopedic Dictionary of the Syriac Heritage:

<place>
  <placeName>Edessa</placeName>
  <desc>
    A city of <placeName>Mesopotamia</placeName>, the capital of the ancient kingdom of <placeName>Osrhoene</placeName>, modern <placeName>Urfa</placeName>.
  </desc>
</place>

The advantage of using TEI, is that it allows Syriaca.org to “markup” or augment the text of a source document with semantic or humanistic interpretive tags. This added markup facilitates searching not just on the raw data of the

---

50 For examples of the variety of scholarly domains for which the TEI Guidelines can be used see Burnard, O’Brien, and Unsworth, <http://www.tei-c.org/About/Archive_new/ETE/Preview/index.xml>, last accessed May 13, 2014.
51 This prose description is based on that of Amir Harrak in Brock et al. 2011, 138–139. It has been reused with permission of Gorgias Press in the Syriaca.org entry for “Edessa,” <http://syriaca.org/place/78>, last accessed May 13, 2014. The XML encoding is simplified and adapted slightly from the raw data at <http://syriaca.org/place/78/tei>.
text (the phrase “Mesopotamia”) but also on encoded interpretations of the text (all “place names” in the text). The power of the computer analysis can then be used to answer increasingly complex queries based on scholarly interpretations, for example a Syriaca.org user might ask for a search result which returned the contents of all <placeName> elements in an article related to a particular person. The power of this sort of data analysis is magnified when one considers the possibility of running such queries over large corpora of texts.

7 Foundational Digital Tools: Linked Open Data

In addition to XML, the other major technological foundation of Syriaca.org is Linked Open Data (LOD). The W3C describes Linked Data, or the Semantic Web of Data, as a plan for “large scale integration of, and reasoning on, data on the Web.” This plan for linking information on the Internet was proposed by Tim Berners-Lee the original architect of the world-wide-web. In Berners-Lee’s vision, “The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.”

For Berners-Lee, three key aspects facilitated linking. The first is the assigning of unique identifiers to data. These unique identifiers are “Uniform Resource Identifiers” or URIs. The linking of related resources occurs when different data sets on the Internet use the same scheme of URIs to link their data together. In the case below, one can see how Syriaca.org uses a URI (“<http://syriaca.org/place/78>”) assigned to represent the place Edessa to allow computer analysis to link together two disparate data points (in this case two different names for the same place):

```xml
<place>
  <placeName ref="http://syriaca.org/place/78">Edessa</placeName>
</desc>
```

---


A city of <placeName ref="http://syriaca.org/place/124">Mesopotamia</placeName>, the capital of the ancient kingdom of <placeName ref="http://syriaca.org/place/145">Osrhoene</placeName>, modern <placeName ref="http://syriaca.org/place/78">Urfa</placeName>.

In this example, the use of a common URI allows the computer to inference that the raw text “Edessa” and “Urfa” are both place names for the same place (<http://syriaca.org/place/78>). Ultimately, a query across multiple documents could compile all of the known names for the place marked as <http://syriaca.org/place/78>. Accordingly, one of the main tasks of Syriaca.org has been to assign URIs to all core entities in its database to enable linking.

A second aspect of Linked Open Data is that the URIs used should also themselves be resolvable Internet addresses. In other words, not only should <http://syriaca.org/place/78> stand for the place “Edessa,” it should be a URL that resolves to a definition of the URI. This document should ideally be available in format or document that computers can read and process without human assistance (“machine readable”). In this case the format is RDF, Resource Description Framework, a format that can be serialized (generated) from other formats, including TEI XML. Syriaca.org has begun to implement this feature. Already the URI <http://syriaca.org/place/78> is also a working URL (web link) which resolves to a human-readable webpage, an entry, “Edessa”, in Syriaca.org’s Syriac Gazeteer: <http://syriaca.org/place/78.html>. An API will also return an instance of the same data in RDF.

A third aspect of Linked Open Data involves making one’s data open and accessible to others. This involves both the use of so-called “open licenses” where copyright does not prevent the sharing and copying of data. As might be self evident, linked data is only of benefit if that data is publically accessible. Accordingly, Syriaca.org has chosen to make its data available under an open license, The Creative Commons Attribution license (CC BY 3.0) which is a

---

55 This example is simplified and adapted slightly from the raw data at <http://syriaca.org/place/78/tei>, last accessed May 13, 2014. It should be noted that this example is in TEI XML. In a strict implementation of linked open data, the information in TEI XML would be first converted to another data format, namely RDF, Resource Description Framework. The standards for RDF are maintained by the W3C at <http://www.w3.org/TR/2014/REC-rdf-schema-20140225/>, last accessed May 13, 2014.

Syriaca.org as a Test Case

This license requires proper scholarly attribution of data in Syriaca.org and prohibits fraudulent use of the data but otherwise allows extensive reuse. Allowing reuse is important both for ensuring the long term preservation of the data (anyone who wishes can download and save a copy) but also for allowing for the possibility that a scholar will transform Syriaca.org's data into something better, beyond the imagination of its original creators. Moreover, the use of open licenses for data reflects a core value of the digital humanities, the democratization of access to knowledge.

Lastly, Berners-Lee’s principles for Linked Open Data emphasize strongly the collaborative nature of linked data. Not only should each project publishing data online create URIs for their own data, they should also “include links to other URIs” (Berners-Lee 2006). In other words, the ultimate goal of Linked Open Data is to begin to build connections between the mass of data on the web through linking of URIs of common entities. In the case of Syriaca.org much of this linking of URIs has been done through linking to either related digital projects in Classics and Medieval History or to larger resources on the Internet. For example, Syriaca.org links its place “Edessa” to both an entry in Pleiades, a scholarly online atlas of ancient world, and to Wikipedia, the largest linked resource on the web.

Benefits to Syriac Studies from XML and Linked Data

XML and Linked Data technologies aid the work of Syriaca.org in two primary ways. The first is perhaps self-evident. Using markup tags and URIs turns the data into a standard form that is easily manipulated or processed in computer applications. The advantage of data standards, however, is not only for machines. There is also a human benefit to shared data standards: the use of common ways to format information facilitates human collaboration and makes it easier for humans to compare information. For example, as we have noted, there are at present perhaps over a hundred different ways to catalogue a Syriac manuscript. This makes comparing manuscript descriptions difficult. By adopting a shared form of communicating information (XML) and a shared

---

57 For more on these licenses see <http://creativecommons.org/>, last accessed May 13, 2014. A discussion of the philosophy behind these licenses can be found in Spiro 2012, <http://dhdebates.gc.cuny.edu/debates/text/13>, last accessed May 13, 2014.
58 As above, this example is actually in TEI XML but would be transformed into RDF to be served as linked open data.
way of identifying important entities (URIs), it becomes easier for scholars to collaborate and share information.\textsuperscript{59}

The ability of digitally enhanced collaboration to dramatically further research has been observed to be one of the major changes wrought by the Internet revolution. (It should be remembered that the World Wide Web was itself was originally conceived of as a way for scientists at the CERN laboratory, Conseil Européen pour la Recherche Nucléaire, to instantly share data and research.) For instance, the now ubiquitous tools of e-mail, social media, and telephony (such as Skype) allow scholars around the world to share their research in minutes rather than days or weeks. Moreover the ability to publish on the web, especially using interactive “Web 2.0” tools such as blogs where readers can comment or micro blogs (such as Twitter) which facilitate real-time conversations now mean that scholars can receive feedback and corrections on their research very quickly. All of these increases in interactivity and speed of communication have sped up the pace of scholarly communication, exchange, and discovery.

Increased speed of dissemination is not the only benefit, however. These new tools for collaboration offer the possibilities for new methods of interactive scholarship. Collective research or publishing projects that would have previously been deemed unfeasible due to the practicalities of geography or communication time can now be undertaken. Following the work of Nielsen, Claire Clivaz has noted several examples of online collaboration in the sciences which humanists might emulate (Clivaz 2013, 159–161). For example the Polymath Project, a blog where collaborators worked together an in only 37 days developed a new proof of a mathematical theorem (27 volunteer collaborators made over 800 comments of over 170,000 words as they worked together on the proof).\textsuperscript{60}

\textsuperscript{59} At the same time, however, it is important to note that while shared standards facilitate data sharing, the same standards can also (unwittingly) stand as institutional barriers excluding scholars who are not conversant in those standards. In particular, it is essential to take steps that digital “literacy” not become a new way to exclude groups who have long been marginalized from scholarship including women, scholars of color, and academics from beyond the Anglophone world. For a reflection on such barriers see Posner 2012, http://miriamposner.com/blog/some-things-to-think-about-before-you-exhort-everyone-to-code/, last accessed November 18, 2014. See the discussion of a related debate about coding and exclusion in Ramsay 2011, http://stephenramsay.us/text/20110108/whos-in-and-whos-out/, last accessed November 19, 2014; and the discussion of Ramsay in Matthew K. Gold, 2012.

\textsuperscript{60} See the account of the Polymath Project in Nielsen 2012, 1–4.
Similar fruits of digital collaboration are also possible in Syriac studies. Since 2012, the Syriaca.org research team has met nearly weekly via Internet tele-conference and used collaboration tools such as Skype, Dropbox, Google Docs and Github to edit research documents in real time. The scholars and research assistants in these meetings have usually been spread over five locations in as many as three countries. The information compiled collectively reveals the power of digital collaboration. For example one team has compiled information on nearly 1,000 Syriac authors, including over 3,000 variants of the authors’ names in Syriac, Arabic, English and other languages. Similarly a group of about half a dozen scholars were able to compile the nearly 2,500 entries in The Syriac Gazetteer (www.syriaca.org/geo) without needing to meet in person (although the team did meet weekly via the Internet).

Not only can digital tools enhance immediate human collaboration on common research, digital standards can make possible unanticipated reuse of data. For example, in the sciences, an increasing number of disciplines are creating shared databanks or repositories that pool together research data. Nielsen and Clivaz have noted that this "data web" not only ensures that new findings in the field are accessible, but it also makes possible new analysis and questions. "As the data web grows, so too will the number and variety of questions that can be asked."61 In other words, by creating large online datasets, Syriaca.org can enable future generations of scholars to re-analyze the knowledge in the field in order to answer questions not yet imagined. This benefit also extends beyond Syriac studies. By publishing its information using the standards of Linked Open Data Syriaca.org is joining a growing community of scholarly projects that are combining their data into a “data web" for research on ancient and medieval history.

Two specific examples of how Syriaca.org plans to share its data are worth noting here. First, Syriaca.org is working with a number of ancient world digital projects as part of New York University’s “Linked Ancient World Data Institute” to add its data to a growing mass of linked data for the Ancient World.62 In particular the Pelagios project for ancient geography has succeeded in linking over “750,000 place references in datasets from 27 partners.”63 Syriaca.org is in the process of submitting the 2,500 entries of the Syriac Gazetteer to this linked data set. Another example is the Medieval Electronic

---

61 Quotation is from Nielsen 2012, 123; see also Clivaz 2013, 160–161.
Scholarly Alliance (MESA), an RDF based search engine at North Carolina State University that currently links together over 100,000 peer-reviewed digital “objects” (such as articles or manuscript descriptions) from 22 federated projects including the British Library, the Walters Art Museum, and the e-codices project.\textsuperscript{64} Syriaca.org is also in the process of applying to include its data in the MESA repository.

We anticipate that the sharing of Linked Open Data will produce a number of benefits. Scholars of ancient and medieval history who are not familiar with Syriac materials will be able to discover relevant data derived from Syriac sources by searching these larger data sets. Similarly, Syriac specialists interested in comparative research will be able to discover research related to Syriac studies through the links between Syriaca.org and the larger data sets. Finally, the analysis of large data sets will most likely suggest new research questions that have yet to be envisioned because the information from Syriac sources has thus far largely been excluded from the corpus of sources used for the history of the Western Mediterranean and Europe.

Conclusions

The experience of Syriaca.org has been that the digital humanities have much of benefit to historians of early Christianity. In terms of the questions posed by this volume, we may offer the following tentative answers: “Do the Digital Humanities have a superficial or deep impact on research and education?” This question is perhaps impossible to answer at present. The digital humanities have only just begun to make their presence felt in Syriac studies. It seems quiet likely that digital approaches and tools will have a long lasting impact, but only time will tell. “Are Digital Tools and Databases just accelerating research, or changing the methods and results?” It is certainly true that digital tools are accelerating research. That change in itself is worthwhile. But the case of Syriaca.org shows the power of digital approaches and standards to change the way that scholars conceptualize their fields, and even more to allow somewhat obscure subfields to be connected and discoverable (through Linked Open Data) to scholars who might not traditionally have used Syriac materials. Lastly, this volume asks, “Do the Digital Humanities offer new solutions to old problems of editing and publishing?” It is to this final question that Syriaca.org provides a resounding yes. The Syriac Gazetteer, our first digital

\textsuperscript{64} See <http://www.mesa-medieval.org/about/>, last accessed May 19, 2014.
publication would never have been feasible as a print publication. Moreover, the two core technical frameworks implemented by Syriaca.org, Extensible Markup Language (XML) and Linked Open Data (LOD), offer new solutions to make the publication of Syriac materials accessible to non-specialists. In short, the experience of Syriaca.org demonstrates some ways in which the emerging tools of the digital humanities offer new paradigms for research and publishing in the history of Christianity and other religions of the ancient and medieval Mediterranean world.

Appendix A. Selected Digital Resources Related to the Study of Syriac

A Comprehensive Bibliography on Syriac Christianity (The Hebrew University of Jerusalem) <http://csc.org.il/db/db.aspx?db=sb>
The Syriac Gazetteer (Syriaca.org) <http://syriaca.org/geo/>
Syri.ac (University of Oklahoma) <http://syri.ac>
The Oliver database (Hill Museum and Manuscript Library) <http://www.hmml.org/olsr.html>
E-ktobe <http://www.mss-syriaques.org/>
Prosopography of the Babylonian Magic Bowls (University of Southampton) <http://www.southampton.ac.uk/vmba/projects/Prosopography%20to%20the%20Babylonian%20Incantations%20Bowls%20An%20Introduction.html>
The Harrak Collection of Iraqi Syriac and Garshuni inscriptions (University of Toronto) <http://www.epigraphy.ca/>
Peshitta Electronic Text Project (Universiteit Leiden) <http://www.hum.leiden.edu/religion/research/research-programmes/antiquity/peshitta-electronic-text-project.html>
Syriac Studies Reference Library (Brigham Young University) <http://lib.byu.edu/collections/syriac-studies-reference-library/>
Syriac Electronic Corpus (Brigham Young University) <http://cpart.maxwellinstitute.byu.edu/home/sec/>
Vatican Syriac Manuscripts (Brigham Young University) <http://cpart.maxwellinstitute.byu.edu/home/vs/>
The Comprehensive Aramaic Lexicon (Hebrew Union College) <http://cali.cn.huc.edu/>
References


“Linked Data.” <http://www.w3.org/standards/semanticweb/data>.


