The Promise and Challenges of Digital Humanities for the Study of Jews and Judaism in Antiquity

Michael L. Satlow

Brown University

In the broad area that falls somewhere under the umbrella-term "biblical studies," there has been an explosion of interest in the digital humanities. There are series (like this one) devoted to it; program units on it at international conferences; countless websites about it; and even degree programs in it.[1] Yet defining exactly what "it" is remains somewhat elusive, and I suspect that there are few scholars who would claim that the early promises of the application of the digital humanities to biblical studies has in any meaningful way transformed the field, or even made much of an impact solving old research problems or opening new ones.

The goal of this paper is to offer a birds-eye, state-of-the-field critical overview of scholarly digital projects dealing with early (post-biblical to Late Antique) Judaism with an eye towards the future. The vast majority of these projects have been geared toward creating data

[1] Journals: Open Theology: https://www.degruyter.com/view/j/opth; Digital Humanities in Biblical, Early Jewish, and Christian Studies, SBL/iSBL (e.g., https://www.eabs.net//EABS/Research-Units/Research_Units/Research_Units_2019/Digital_Humanities_in_Biblical_Studies.aspx); MA program for Eep Talstra Centre for Bible and Computer: http://www.godgeleerdheid.vu.nl/en/research/institutes-and-centres/eep-talstra-centre-for-bible-and-computer/index.aspx.

archives that provide accessibility and searching capabilities. This is no small thing, having changed the ease and speed with which scholars do their work. At the same time, though, I want to argue that a more careful consideration of data modeling, infrastructure, and interface, could make these, and other sites, much more likely to contribute to having a transformative impact on the field.

The paper has two parts. In the first, I will discuss best practices behind the creation of data, specifically with reference to existing projects. In the second, I will turn mainly to projects outside of biblical studies that focus on analysis and argument as a way to reflect on the unrealized potential of applying digital techniques to the data of biblical studies.

1. Data

In 2016, a group of scholars responded to what they saw as the need to articulate a framework that would facilitate the ability of machines to find, analyze, and reuse scientific data automatically with what they termed the "FAIR Principles." The four components of these principles are: Findability, Accessibility, Interoperability, and Reusability.[2] Their application to the sciences is readily apparent. Scientists frequently deal with reams of data that are, in various ways, not easily compatible. Finding, harvesting, transforming, and analyzing datasets can be done manually but only at significant labor and cost. The idea behind FAIR is that the

[2] M. Wilkinson, M. Dumontier, I. Aalbersberg, *et al.*, "The FAIR Guiding Principles for Scientific Data Management and Stewardship," *Sci Data* 3, 160018 (2016) doi:10.1038/sdata.2016.18. The current version of the FAIR Principles can be found at: https://www.go-fair.org/fair-principles/.

application of these four principles to data would significantly aid the entire scientific community.

Scholars in the humanities rarely, if ever, face "big data" at anywhere near the level of scientists. Yet the problem that digital humanists face – not just in biblical studies – is similar. There are many research questions that involve the machine analysis of large amounts of data but creating such massive datasets is beyond the means of almost any individual scholar on their own. For any single dataset in the humanities to be maximally useful, it too should adhere to the FAIR principles.

Below, using the FAIR rubrics (although, for organizational reasons, not in the order of the acronym), I discuss the current landscape of major data silos relating to early Judaism. The goal of this discussion is not to critique these projects, all of which are fantastic resources and most of which were developed long before the vision of analysis over multiple datasets in the humanities was even on the horizon. Rather, it is to highlight general issues to which project directors, institutions, and the entire scholarly community might attend in order to further drive us toward realizing some of the promises of the digital humanities, which I discuss in the second part of this essay.

1.1 Accessibility

Accessibility is the most basic principle. Obviously, scholars cannot analyze data they do not have access to it. There are three issues that generally fall under this rubric: the degree to which a dataset is open; the ease of automatic accessibility; and the accessibility of the metadata.

The datasets relating to the study of Jews and Judaism in antiquity generally fall into three levels of openness. The first, and most restrictive, are the databases that require payment. Some databases like *Accordance*, the *Bar-Ilan Responsa Project*, and *The Sol and Evelyn*

Henkind Talmud Text Databank, all limit access to their data to those who buy or license their products and restrict the ability of those users to share and reuse (see 1.3, below) their data.[3] Several databases that are ancillary but relevant to scholars of early Jews and Judaism have a similarly limited accessibility. The *Thesaurus Linguae Graecae*, a huge corpus of Greek literature, needs a paid subscription for access as does *Sources Chrétiennes Online*, which is just being released.[4]

On the other end of the spectrum of accessibility are those databases that use a protocol that is open, free, and universally implementable. Two comprehensive collections of rabbinic texts have these qualities, *Mechon-Mamre* and *Sefaria*.[5] Other, more limited corpora of rabbinic texts also fall into this category: *Corpus Tannaiticum*, *Tosefta Online*, *Yerushalmi Online*, *Digital Mishnah*, and the *THALES Lectionary Project*.[6] Several sites dealing the material culture of Jews in antiquity also maintain open access. These include: *Inscriptions of Israel/Palestine*,

[3] *Accordance*: https://www.accordancebible.com/; the *Bar-Ilan Responsa Project*: http://responsa-forum.co.il/www/?page_id=322&lang=en; *The Saul and Evelyn Henkind Talmud Text Databank*: Saul Lieberman Institute of Talmudic Research of the Jewish Theological Seminary of America, *Ma'agar 'ede nusah shel ha-Talmud ha-Bavli* (2011).

[4] *Sources Chrétiennes Online*: https://about.brepolis.net/sources-chretiennes-online-sco/.

[5] *Mechon-Mamre*: https://www.mechon-mamre.org/; *Sefaria*: https://www.sefaria.org/.

[6] *Corpus Tannaiticum*: http://mishna.huma-num.fr/; *Tosefta Online*: https://www.toseftaonline.org/; *Yerushalmi Online* https://www.yerushalmionline.org/; *Digital Mishnah*: https://www.digitalmishnah.org/, and the *THALES Lectionary Project*: http://www.lectionary.eu/.

The Levantine Ceramics Project, and the National Treasures Online of the Israel Antiquities Authority.[7] There are also several useful tools that are freely available, especially *Ma'agarim* (the Historical Dictionary Project of the Academy of the Hebrew Language), the *Comprehensive Aramaic Lexicon*, and Jastrow Online.[8] Many early secondary texts that have fallen out of copyright are also freely available at Google Books and HebrewBooks.[9]

Several other databases that are relevant to scholars of early Jews and Judaism are also fully open. These include *The Perseus Digital Library*, the *Digital Latin Library*, *Coptic Scriptorium*, *Papyri.info*, *Syriaca.org*, and several epigraphic databases.[10]

---

[7] *Inscriptions of Israel/Palestine*: https://library.brown.edu/iip/index/, *The Levantine Ceramics Project*: https://www.levantineceramics.org/; and the National Treasures Online of the Israel Antiquities Authority: http://www.antiquities.org.il/t/default_en.aspx.

[8] *Ma'agarim: The Historical Dictionary of the Hebrew Language*: (the Historical Dictionary Project of the Academy of the Hebrew Language): https://maagarim.hebrew-academy.org.il/Pages/PMain.aspx; *Comprehensive Aramaic Lexicon*: http://cal.huc.edu/; and *Jastrow Online*: http://www.tyndalearchive.com/TABS/Jastrow//.

[9] *Google Books*: https://books.google.com/; HebrewBooks: https://hebrewbooks.org/.

[10] *The Perseus Digital Library*: http://www.perseus.tufts.edu/hopper/; the *Digital Latin Library*: https://digitallatin.org/; *Coptic Scriptorium*: https://copticscriptorium.org/; *Papyri.info*: http://papyri.info/; *Syriaca.org*: http://syriaca.org/. For epigraphical databases, see, for example, *Epigraphic Database Heidelberg*: https://edh-www.adw.uni-heidelberg.de/home; *Epigraphic Database Bari*: http://www.edb.uniba.it/. See also the database of the American Numismatic Society: http://numismatics.org/search/.

There is a third set of databases that are free but require registration for access. The principle databases of this type require registration with the Friedberg Jewish Manuscript Society.[11]

To this landscape of mixed access one might add the more complex and confusing issues of both automated access and accessibility of metadata. For certain kinds of projects in digital humanities it is important, and sometimes crucial, to be able to access selected data from different sites in real time. This requires that the data be hosted within an infrastructure that contains an Application Programming Interface (API). Most of the projects listed above do not seem to have such capability (*Sefaria* is a welcome exception), or if they do, they are not transparent about how to use it. Similarly, for most of these projects, the state of and access to the metadata (information about the data) is, at minimum, unclear. This absence makes the database less useful for projects that are based on analysis of metadata.

Two examples might help to illustrate the problem more concretely. The *Digital Mishnah* project seeks to create a digital edition of the Mishnah, along the lines of some of the emerging best practices in thinking about how digital editions are to be constructed. Much of the data that might go into such an edition – e.g., manuscript transcriptions and images, word parsing, secondary references – already exist digitally. The lack of accessibility, however, means that the project will need to enter into complicated negotiations with several partners, or redundantly recreate the digital data.

[11] Friedberg Jewish Manuscript Society: https://fjms.genizah.org/.

A second, more hypothetical, example is metadata analysis. Metadata (such as library catalogues) can be analyzed on their own.[12] One can imagine metadata analyses being done on manuscripts of rabbinic texts (e.g., an analysis of the size of the pages of hundreds of manuscripts in different locations) or the Cairo Geniza documents (for which extensive digital metadata exists). In these cases, however, accessibility stands as a barrier.

1.2 Findability

For data to be maximally findable, it should have four features: it should have a unique and persistent identifier; be described with rich metadata; the metadata should include the identifier of the data they describe; and the data and metadata should be registered in a searchable resource. The absence of one or more of these features in a dataset makes it much harder for a machine, in particular, to automatically locate and use the data.

Very few of the databases relating to the study of Jews and Judaism contain all of these features, and some contain none. The absence of attention to these features is understandable. Most of these databases were built to be discovered and used by humans and, for visibility, have relied on traditional marketing tactics used for scholarly projects, such as announcements on listservs. Since there is also a relatively small universe of sites with data relevant to this field of study, the developers of the sites saw little need to expend the resources necessary to make the data findable to machines.

The lack of conformity to these standards, though, has the potential to limit the use of the data. One can now begin to imagine broad research projects that combine data in this field with

[12] For one application of such an analysis, see, *Footprints: Jewish Books through Time and Place*: https://footprints.ctl.columbia.edu/.

data (e.g., textual, material) produced and curated by scholars in other specializations. Making one's data findable would help such projects locate it.

When it comes to the digital humanities, both the determination of the metadata and the assignment of unique and persistent identifiers to data are related and raise complex issues. The determination of what extrinsic information goes in the metadata to make it useful is largely a decision of a specific scholarly community. Some features are rather intuitive, such as latitude and longitude coordinates for data that deal with places, but others (e.g., protein names) are determined, and given authoritative spellings or ids, within a specific community of practice.

The Textual Encoding Initiative (TEI), probably the most widely used standard for texts in the digital humanities, provides standard ways for encoding a very wide variety of metadata, but different projects include different elements of that metadata. Many projects straddle digital communities of practice, which makes the determination of metadata standards quite difficult. *Sefaria,* for example (which to my knowledge does not encode metadata in TEI format), contains rabbinic texts; texts from different places and time periods; and texts in different languages. In fact, there are many other scholarly communities that may be interested in these texts as well. For the communities to find these texts, though, there must be metadata that uses expected terms. One challenge that these different communities face is determining the information that metadata should contain and the terms used to represent them.

The issue of metadata is related to the issue of identifiers, or URIs (universal resource identifiers). In theory, this is a simple concept: each datum needs its own unique, persistent, id. In *Inscriptions of Israel/Palestine,* for example, each inscription has its own id, as do entries in the *Levantine Ceramics Project*. The practical problem, though, comes especially when encoding texts. What counts as needing a unique URI? Is it a whole text (e.g., the Mishnah), a

pericope (e.g., Mishnah Berakot 1:1), or a word within a pericope? The answer to that question depends very much on how the encoder envisions the data will be used.

Stable URIs for data are basic to the use of that data via accepted API protocols. An emerging standard for images is the International Image Interoperability Framework (IIIF).[13] The two standard API protocols are the Canonical Text Service (CTS) protocol, which is now being replaced with the more flexible Distributed Text Services (DTS) protocol. The latter especially supports the ability to retrieve texts at any level of a citation hierarchy (as long, of course, as a citation hierarchy actually exists).[14] For data to work with any of these protocols, however, they must be encoded and identified with consistently formed URIs.

1.3 Reusability

Data reusability is a bit of an odd idea in the humanities. Humanistic disciplines have a long-established set of practices dealing with the use and citation of both primary and secondary sources. It is a little harder for many humanists to think of how this data – textual, material, images – can and should be automatically machine-processed, and how "credit" is assigned in such cases.

The question of reusability in the context of the digital humanities is a big one and well beyond the scope of this paper. I think, though, that we are all ultimately interested in contributing to the production of knowledge by optimizing reuse of our data, whatever we

[13] See, https://iiif.io/.

[14] On CTS, which was developed for the *Homer Multitext* project, see, https://wiki.digitalclassicist.org/Canonical_Text_Services. On DTS: https://distributed-text-services.github.io/specifications/.

understand that to be. In this respect, it is worth briefly discussing two aspects of data encoding that touch upon reusability, metadata (again) and licensing.

For data to be reusable its scope and attributes should be richly defined in its metadata. If, for example, we want to find all examples of a certain kind of ancient pot across all museum databases, we might write a program to automate this process. The program can only work, though, if it can identify the relevant terms in the metadata. With reams of relevant data coming online each year, the issue of identification and rich description takes on increasing importance.

One of the most important issues related to reusability is licensing. This issue relates back to that of accessibility. The more restrictions on the data, the less reusable it is. Best practices in digital humanities today call for Creative Commons licensing, with the exact license embedded in the metadata. A Creative Commons license has the advantage of being machine-readable, so an automated process can discern what it can do, and not do, with data without human intervention.[15] The most accommodating of licenses, and the one that will make data most accessible, is the CC BY 4.0, which allows users to reuse and remix data, for noncommercial and commercial purposes, but with attribution and notice if changes were made.

The fundamental idea behind reusability is that one never knows to what use, now and in the future, a scholar might put one's data. Projects that build into their data the infrastructure that will allow reuse are far more likely to make an impact than projects that don't.

1.4 Interoperability

Merging datasets by hand can be resource-intensive, often to the extent that it puts certain research projects beyond the limits of most scholars in the humanities, who have access to

[15] Creative Commons: https://creativecommons.org/.

limited resources (i.e., time and money). Data that are produced according to the standards of a particular scholarly community of practice (but also with an eye towards other communities that are more likely to find them useful) will be far more useful to these scholars.

Producing data according to the standards of a scholarly community requires, at minimum, two things. First, it requires a certain uniformity of encoding. Texts, for example, that are structured according to TEI guidelines are far more useful to scholars than texts that are not. Second, the data, and especially the metadata, should use common vocabularies or authority lists. To denote, for example, a reference to a place there should be a shared way of describing that place, even if the actual reference may not use that precise term of identification. Authority lists are not unique to digital humanities; libraries in particular have spent much time developing them to assure, for example, that the works of a single author are correctly attributed, even when there are two or more authors of the same name.

Standard vocabularies are the backbone of Linked Open Data (LOD).[16] The goal of LOD is to allow automated processes to harvest data across projects according to certain predetermined criteria. *Pleiades*, for example, assigns unique identifiers to places.[17] Data that use that unique identifier as part of their metadata can easily be identified, linked to each other, and processed. There are a wide variety of potential vocabularies, only some of which are in development. *Inscriptions of Israel/Palestine,* for example, uses *Pleiades* ids for places; *Getty*

[16] For an introduction to LOD, see Jonathan Blaney, "Introduction to the Principles of Linked Open Data," *The Programming Historian* (2019): https://programminghistorian.org/en/lessons/intro-to-linked-data.

[17] *Pleiades*: http://pleiades.stoa.org.

Thesaurus ids for objects; and *PeriodO* ids for time-periods.[18] There remains, though, no authority list for Rabbis. *Syriaca.org* provides unique ids for persons in their corpus, but I know of no similar list for bishops in Greek and Latin sources.

1.5 Reflections

There is a good amount of digital data relating to the study of Jews and Judaism in antiquity, although there remain gaps. Availability of high-quality transcriptions of manuscripts and important editions of rabbinic texts, for example, remains spotty. The largest lacuna, though, seems to me to be in the material evidence; much of the archaeological data relating to Jews and Judaism in antiquity, both in and out of the State of Israel, is not in an accessible and usable digital form. Much of the raw archaeological digital data are destroyed or privately shelved after the print publications of the final reports. Digitizing the records and stores of the Israel Antiquities Authority alone is a momentous task. There remains work to be done.

The larger issue, though, that I hoped to have highlighted in this survey is that the data that do exist most frequently do not exist in a form that makes them maximally useful for projects in the digital humanities. There are digital texts, but many of them are not easily accessible by machine; some are in proprietary formats; and almost none are structured according to an accepted standard with rich (and linked) metadata. These databases are of great value to scholars seeking quick access and simple searching, but have more limited utility for analysis.

[18] *Getty Art and Architecture Thesaurus*: https://www.getty.edu/research/tools/vocabularies/aat/index.html; *PeriodO*: http://perio.do/en/.

The goal of opening and encoding data for maximal use is certainly laudatory, but it would be misleading not to mention the costs and challenges of doing so. Data encoding, of course, takes resources. Some projects understandably raise the money that they need to continue their operations through sales and subscriptions (e.g., *Bar-Ilan Responsa Project*; *TLG*); some are institutionally and grant supported (e.g., *Pleiades*); others rely on private benefactors (e.g., *Sefaria*; Friedberg Jewish Manuscript Society). It takes time and money to add rich metadata and standard, linked vocabularies to data. This is all the more so when the standardized, linked vocabularies are still being developed by scholars.

The larger challenge, then, is organizational and administrative. Most of these projects have a hierarchical and controlled administrative structure. There is often good reason for this; it is otherwise difficult to assure consistent and efficient encoding. Yet in the end, the sustainability of such a structure – and more troubling, of the data and the interfaces that have been built to deliver it – is intrinsically shaky. Unless the data are deposited in an institutional repository (as in *Inscriptions of Israel/Palestine*), it is unclear how some of these projects will make accessible, not to mention enhance, their data once their funding ends or their project directors move on.

This is a widespread issue, not one confined to the study of Jews and Judaism in antiquity. While we need to develop different models of sustainability, some form of open access and distribution seems promising.[19] One might, for example, deposit one's data in Github

[19] There is increasing discussion of sustainability planning. See, for example, *Sustaining Digital Resources: A Briefing Paper for Leaders of Projects with Scholarly Content*:

where somebody else can download it, enhance it (e.g., add linked ids), and upload it again for free reuse.[20] One challenge to implementing such a model, though, is the ability to develop a strong enough network of scholars interested in developing standards, reusing, and building on each other's data. This, in turn, depends to some extent on the institutional recognition and rewards given for such work.

Given the difficulties in preparing good digital data, we might well ask, Why bother? What might we be able to do with good digital data that we cannot do now?

2. Analysis

In this section, I turn from the data to consider things that we might want to do with properly structured data. We take access and searching capability for granted now, but there are many more things that we can do with good data. Here I draw from other digital projects primarily outside of the study of Jews and Judaism in antiquity to chart some possible directions for the application of digital humanities in this field.

2.1 Textual Editions

What does it mean to create a text edition in the twenty-first century? Can it, or should it, mean something different from what we have always done? A few projects in the field, such as *Digital Mishnah*, *Corpus Tannaiticum*, and *THALES*. These projects are excellent demonstration projects that begin to tap the potential of digital editions, using especially tools that create

https://sr.ithaka.org/wp-content/uploads/2015/08/BRIEFING_PAPER-for-digital-project-leaders.pdf.

[20] Github: https://github.com/.

synopses of text transcriptions, translations, and images. More ambitious, the *Open Greek and Latin Project* explores how a born-digital commentary can be intrinsically different from the traditional print version, not modeled on it.[21]

The greatest potential of digital editions of ancient texts is their ability to integrate commentaries from different authors while sustaining and preserving scholarly discussions. A platform and infrastructure that allows scholars to participate in the activity of commentary in a way that preserves their arguments (and support for them), and that can grow over time, is potentially much more useful than the traditional commentary that reflects the thinking of just a single scholar or a small group. Commentary can be dynamic rather than static.

There are different models for thinking about how such a text edition can be implemented, but it is clear that some kind of Canonical or Distributed Text Services model can enhance the possibilities. Different scholars can reference and comment on the same fragment of a text – using the same ID – even in different platforms and their comments can be easily brought together. Ultimately, it may be possible for a machine to collect automatically and show to users everything said about a particular passage. This model differs from a dynamic and organically developed commentary that then gets preserved within the data itself.

2.2 Bringing Data Together

Data amalgamation can occur in many different ways. A robust data ecosystem, with linked open data, allows scholars to bring and see together different kinds of potentially relevant data from all over the Web. *Perseus* has long used this model for data internal to one server, but

[21] *Open Greek and Latin Project*: https://www.dh.uni-leipzig.de/wo/projects/open-greek-and-latin-project/.

Pleiades and the *Pelagios Network* are actively experimenting with this model for data that originates from many sites.[22]

Juxtapositions, particularly unusual ones, can tell stories and spark new ones. Being easily find and assemble a package of data of different kinds from across multiple databases according to a range of sophisticated criteria can spur new knowledge. The creation of these juxtapositions, though, is largely dependent on the ability of automated processes to find materials across different databases.

2.3 Topic Modeling

Topic modeling is primarily a text-mining tool that unearths things within a text – words, sentences, or other semantic units primarily – that tend to occur together. These correlations can reveal underlying connections within a single text or corpus while also tracing correlations over time. One of the most interesting applications of this technology is *Mining the* Dispatch, which uses topic modelling to trace "topics" in runaway slave ads from the 1860-1865 in the Richmond *Daily Dispatch* as a way to contribute to slave behavior during the Civil War (and its representation by owners). Usually, software packets such as MALLET are used for such analyses.[23]

Topic modeling has particularly interesting applications for dealing with ancient texts, particularly for answering questions that deal with modern scholarly definitions. One might, for example, investigate the clusters of words within which certain ancient words that we commonly translate as "religion" or "piety" appear, thus helping to reveal the larger concepts behind those

[22] *Pelagios Network*: https://pelagios.org/.

[23] MALLET: http://mallet.cs.umass.edu/topics.php.

words, and the extent to which they correspond or not to the modern words we use to translate them.[24]

2.4 Historical Simulations

Historical narratives are, or are based on, models of how people have done things. To my mind, one of the more intriguing applications of computing to historical data has been quantified into interactive ways of presenting these models. This kind of work was pioneered by the Cambridge Group for the History of Population and Social Structure.[25] A example of such a project that is tangentially relevant to our topic is *ORBIS: The Stanford Geospatial Network Model of the Roman World*, which models travel and communication networks through the Roman Empire.[26] The same arguments about the interrelationship between distance, topography, travel routes, political borders, and cost could be made in a narrative form, but probably less effectively. All kinds of ancient social interactions (e.g., demography) can be modeled in a

[24] The question of whether there was "religion" in antiquity has been most powerfully raised in Brent Nongbri, *Before Religion: A History of a Modern Concept* (New Haven: Yale University Press, 2013); Carlin A. Barton and Daniel Boyarin, *Imagine No Religion: How Modern Abstractions Hide Ancient Realities* (New York: Fordham University Press, 2016); and Nickolas P. Roubekas, ed., *Theorizing "Religion" in Antiquity* (Sheffield: Equinox, 2019).

[25] Cambridge Group for the History of Population and Social Structure: https://www.campop.geog.cam.ac.uk/.

[26] *ORBIS: The Stanford Geospatial Network Model of the Roman World*: http://orbis.stanford.edu/.

similar fashion, and can become particularly effective when data from several different silos are analyzed together.[27]

2.5 Lexicography and Linguistics

There are a variety of lexicographical and linguistic analyses that can be done from the combined digital texts produced in antiquity. Natural Language Processing – using parsers such as the Classical Language Toolkit – can be developed and refined to automate the parsing of individual words.[28] Parsing entire corpora, in turn, facilitates not only searching but also other kinds of analyses such as word-counting and other kinds of visualizations. There is much work of this type to be done in free and publicly accessible databases.[29]

There are other kinds of more complex lexical and semantic analyses that can be performed over the range of Jewish writings from antiquity. Particularly intriguing are the possibilities of automated analyses of both intertextuality and rhyme and rhythm in ancient poetry.

---

[27] A non-digital but quantitative model for what can be done with demography in the ancient world (and scaled up with digital data) is Roger S. Bagnall and Bruce W. Frier, *The Demography of Roman Egypt* (Cambridge: Cambridge University Press, 1994).

[28] Classical Language Toolkit: http://cltk.org/. *Coptic Scriptorum* is using NLP effectively for their data.

[29] *The Global Philology Project* (http://www.dh.uni-leipzig.de/wo/projects/global-philology-project/) is an ambitious attempt to consider what can be accomplished by making vast textual data in many languages machine-readable.

2.6. Visualization

Visualization really isn't a topic in and of itself, but it is important to say something generally about it. Computers are exceptionally good at visualizing data. One of the more obvious applications of visualization is geographical. Actually seeing distributions and densities can play a valuable role in driving new questions and constructing arguments.

Visualization is an important feature of many text-mining software packages.[30] Even a simple visualization like a word cloud can quickly reveal the central concerns of a text or, more interestingly, a set of texts selected for certain surface similarities (e.g., date and location).

2.7 Network Analysis

There has been a lot of recent historical work that utilizes network analysis. Network analysis statistically models the relationship between, among other things, objects, texts, places, or people. Analyzing the relationship between people or groups of people (who become the nodes of analysis, with the connections between them known as edges) is the domain of Social Network Analysis (SNA). SNA is most commonly used for understanding connections between people using social media, where large amounts of data can be used, or for improving the flow of information within organizations such as corporations.

Digital humanists have only recently begun applying SNA methods to historical data. Off-the-shelf, open-source software packages such as Gephi have lowered the bar for historians to model their data and thus reveal new information about how knowledge was transmitted between people, schools, and groups.[31] Two projects, *Mapping the Republic of Letters* and *Six*

[30] See, for example, Voyant Tools: https://voyant-tools.org/.

[31] Gephi: https://gephi.org/.

Degrees of Francis Bacon, model the potential for the use of SNA for understanding the flow of knowledge between intellectuals.[32] *Footprints: Jewish Books Through Time and Place* also is beginning to use this technology.

I see two potentially interesting applications of SNA to the study of Jews and Judaism in antiquity (although there are, of course, many more). The first focuses on bibliography and citations. SNA is frequently used now to map citations in academic articles, thus allowing us to visualize specific discursive webs of scholars.[33] The existence of different "schools" of scholarship in the study of ancient Judaism has been discussed, but the data used has been qualitative.[34] By focusing on who cites whom, SNA offers a way to chart more precisely how these schools are constituted and how knowledge moves between them.

A second potential application is to better understand the relationship between the intelligentsia in antiquity. With a colleague, Michael Sperling, I am now working on such a project in which we analyze the citation chains (e.g., Rabbi X says in the name of Rabbi Y…)

---

[32] *Mapping the Republic of Letters*: http://republicofletters.stanford.edu/; *Six Degrees of Francis Bacon*: http://www.sixdegreesoffrancisbacon.com/?ids=10000473&min_confidence=60&type=network.

[33] Citation analysis falls under the broad category of bibliometrics. See, Silvia Salini, "An Introduction to Bibliometrics," in Tony Greenfield and Sue Greener, eds., *Research Methods for Postgraduates* (3rd edition; Chichester: John Wiley & Sons, 2016), 130-143.

[34] Cf. Steven Fine, *Art, History and the Historiography of Judaism in Roman Antiquity* (Leiden: Brill, 2014).

found in the Babylonian Talmud as constituting a network.[35] We will soon report our preliminary results (and publicly release all of our data and software tools) but we can already see how such an analysis can contribute, uniquely, to how we envision both the actual social network of the rabbis as well as the citation and redaction practices of the editors of the Babylonian Talmud.

3. Conclusions

This paper is meant to provide a broad overview of where we are in terms of applying the digital humanities to the study of Jews and Judaism in antiquity, and where we might go. While I see many promising applications for digital techniques, there remain challenges. I have highlighted here challenges relating to data. While more data than ever are available, with more continuously being added, the way that those data have been released do not always make them easy to use. Creating data and metadata in a clean, accessible, and usable form is not always easy ,but attention to the details of doing so are critical for driving the field forward.

I have mentioned but not dwelled here on two other challenges, resources and institutional credit, because they will largely be familiar to the readers of this journal. That does not mean that they are insignificant. Creating and cleaning good data takes time and often money, and its institutional pay-off is uncertain. For scholars to become more comfortable investing their resources in creating good data and then sharing it for reuse, other scholars must first recognize their contributions as true "scholarship," and be willing to credit them accordingly

[35] See, Joshua Waxman, "A Graph Database of Scholastic Relationships in the Babylonian Talmud," https://dev.clariah.nl/files/dh2019/boa/0229.html.

– only then will the institutional organs that grant tenure and push for other metrics be more willing to "credit" such activities. The virtuous circle that encourages the production of good data that lead to good digital analyses that lead to new knowledge begins at home, with us recognizing the genuine scholarly work of data production and dissemination.

Works Cited

Accordance. https://www.accordancebible.com/.

American Numismatic Society: http://numismatics.org/search/.

Bagnall, Roger S. and Bruce W. Frier, *The Demography of Roman Egypt*. Cambridge: Cambridge University Press, 1994.

Bar-Ilan Responsa Project: http://responsa-forum.co.il/www/?page_id=322&lang=en.

Barton, Carlin A. and Daniel Boyarin, *Imagine No Religion: How Modern Abstractions Hide Ancient Realities*. New York: Fordham University Press, 2016.

Blaney, Jonathan. "Introduction to the Principles of Linked Open Data," *The Programming Historian* (2019): https://programminghistorian.org/en/lessons/intro-to-linked-data.

Cambridge Group for the History of Population and Social Structure: https://www.campop.geog.cam.ac.uk/.

Canonical Text Services: https://wiki.digitalclassicist.org/Canonical_Text_Services.

Classical Language Toolkit: http://cltk.org/.

Comprehensive Aramaic Lexicon: http://cal.huc.edu/.

Coptic Scriptorium: http://copticscriptorium.org/.

Creative Commons: https://creativecommons.org/.

Digital Latin Library: https://digitallatin.org/.

Distributed Text Services: https://distributed-text-services.github.io/specifications/

Epidoc: https://sourceforge.net/p/epidoc/wiki/Home/.

Epigraphic Database Heidelberg: https://edh-www.adw.uni-heidelberg.de/home.

Epigraphic Database Bari: http://www.edb.uniba.it/.

FAIR Principles: https://www.go-fair.org/fair-principles/.

Fine, Steven. *Art, History and the Historiography of Judaism in Roman Antiquity*. Leiden: Brill, 2014.

Footprints: Jewish Books through Time and Place: https://footprints.ctl.columbia.edu/.

Friedberg Jewish Manuscript Society: https://fjms.genizah.org/.

Gephi: https://gephi.org/.

Getty Art and Architecture Thesaurus:

https://www.getty.edu/research/tools/vocabularies/aat/index.html.

HebrewBooks: https://hebrewbooks.org/.

Inscriptions of Israel/Palestine: http://www.brown.edu/iip.

International Image Interperoperability Framework: http://www.iiif.io.

Israel Antiquities Authority (National Treasures Online):

http://www.antiquities.org.il/t/PeriodsList_en.aspx.

Jastrow Online: http://www.dukhrana.com/lexicon/Jastrow/.

*Levantine Ceramics Projec*t: https://www.levantineceramics.org/.

Ma'agarim: http://maagarim.hebrew-academy.org.il/Pages/PMain.aspx.

MALLETT: http://mallet.cs.umass.edu/topics.php.

Mapping the Republic of Letters: http://republicofletters.stanford.edu/.

Mechon Mamre: https://www.mechon-mamre.org/.

Mining the Dispatch: http://dsl.richmond.edu/dispatch/Topics.

Nongbri, Brent. *Before Religion: A History of a Modern Concept*. New Haven: Yale University Press, 2013.

Open Greek and Latin Project: https://www.dh.uni-leipzig.de/wo/projects/open-greek-and-latin-project/.

Open Theology: https://www.degruyter.com/view/j/opth.

ORBIS: The Stanford Geospatial Network Model of the Roman World: http://orbis.stanford.edu.

Papyri.info: http://papyri.info/.

Pelagios Network: https://pelagios.org/.

PeriodO: http://perio.do/en/.

Perseus Digital Library: http://www.perseus.tufts.edu/hopper/.

Pleiades: https://pleiades.stoa.org/.

Roubekas, Nickolas P., ed., *Theorizing "Religion" in Antiquity*. Sheffield: Equinox, 2019.

Salini, Silvia. "An Introduction to Bibliometrics," in Tony Greenfield and Sue Greener, eds., *Research Methods for Postgraduates*. 3rd edition; Chichester: John Wiley & Sons, 2016, 130-143.

Sefaria: https://www.sefaria.org/texts.

Six Degrees of Francis Bacon: http://www.sixdegreesoffrancisbacon.com/

The Sol and Evelyn Henkind Talmud Text Databank: Saul Lieberman Institute of Talmudic Research of the Jewish Theological Seminary of America, *Ma'agar 'ede nusah shel ha-Talmud ha-Bavli* (2011).

Sources Chrétiennes Online: https://about.brepolis.net/sources-chretiennes-online-sco/.

Sustaining Digital Resources: A Briefing Paper for Leaders of Projects with Scholarly Content: https://sr.ithaka.org/wp-content/uploads/2015/08/BRIEFING_PAPER-for-digital-project-leaders.pdf.

Syriaca.org: http://syriaca.org/.

Text Encoding Initiative: http://www.tei-c.org/.

THALES Lectionary Project: http://www.lectionary.eu/.

Thesaurus Linguae Graecae: http://stephanus.tlg.uci.edu.

Tosefta Online: http://www.toseftaonline.org/.

Voyant: https://voyant-tools.org/.

Waxman, Joshua. "A Graph Database of Scholastic Relationships in the Babylonian Talmud," https://dev.clariah.nl/files/dh2019/boa/0229.html.

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* "The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci Data* **3,** 160018 (2016): doi:10.1038/sdata.2016.18.

Yerushalmi Online: http://www.yerushalmionline.org/.