SUMMARY OF PROCEEDINGS OF THE "LINKING THE MIDDLE AGES" WORKSHOP (MAY 11-12, 2015) at the University of Texas at Austin[1]

Authors: E. Turnator, A. Bolintineanu, T. Rose-Steel, B. Whearty, M. Widner
Editors: M. Davis, M. Esteva, J. Hecker, T. Kealy, L. Kopár, K. Norsworthy, J. Reeves Flores
Contributors: A. Brin, L. Isaksen, E. Kansa, A. Rabinowitz.

[1] All links cited in this work were accessible by June 30, 2015.

CONTENTS

Introduction……..1

Workshop Organization and Content……..2

Summary of Day 1 Discussions.....3

Common Discussion Points……6

Summary of Day 2 Discussions.....13

Research-Related Data Group.....13

Teaching-Related Data Group..... 21

Peer-Review and Publications-Related Data Group.....29

Workshop Participants.....34

APPENDIX……..42

Workshop Program……..43

Presentation Abstracts……..47

Linked Open Data Overview…… 51

Workshop Survey Questions……..57

INTRODUCTION

In May 2015 the CLIR/Mellon Foundation Postdoctoral Fellows in Medieval Data Curation convened a two-day workshop on the sharing and publishing of linked open data (LOD). Funded by a CLIR/Mellon microgrant, the workshop brought together librarians, technologists, and scholars to exchange ideas on the challenges posed to medievalists in sharing data on digital platforms. More than thirty experts took part in the workshop, which was held at The University of Texas at Austin on May 11-12, 2015. Participants presented their work in linked open data operational sites and took part in discussions about the obstacles and opportunities of LOD in three areas: research, teaching, and publication. This paper documents the possibilities and problems of applying LOD in medieval studies. It focuses specifically on its application to the field's data, which requires medievalists to work together with librarians and technologists, and considers a technical infrastructure that can maintain and proliferate this type of collaborative work.

Recent developments in digital technologies and their heretofore external impact on research, teaching, and publication in the field of medieval studies provided the backdrop for our workshop. The direct impact of the Modern Language Association's (MLA) "Guidelines for Evaluating Work in Digital Humanities and Digital Media" and, more recently, the American Historical Association's (AHA) "Draft Guidelines for the Professional Evaluation of Digital Scholarship in History"[2] have spurred the necessity of our conversations. In addition to the recommendations made by academic associations for evaluating the quality of scholarly digital work, the fact that federal and international agencies now require digital project proposals to have a long term data management plan also impacts the direction of digital scholarship in humanities. Evaluation of digital scholarship directly affects research, teaching, and publishing in the humanities. The "Linking the Middle Ages" workshop addressed all three areas.

Making data available as LOD, or providing robust metadata in general, requires significant collective effort and allocation of both time and resources on the part of researchers, librarians, and technologists. It is expensive to discover, create, and

[2] MLA, "Guidelines for Evaluating Work in Digital Humanities and Digital Media", https://www.mla.org/guidelines_evaluation_digital. This document was created in 2000 and revised in 2012; AHA, "Draft Guidelines for the Professional Evaluation of Digital Scholarship in History" http://blog.historians.org/2015/04/draft-guidelines-evaluation-digital-scholarship/. This document was published in May 2015.

maintain linked data. We need, therefore, to articulate the value of this work and we must evaluate whether the effort is justified. With that in mind, this workshop focused on providing a blueprint for future medieval LOD projects, each of which will need to develop its own approach to metadata and decide on its own level(s) of granularity.

WORKSHOP ORGANIZATION AND CONTENT

Workshop Program (See Appendix)

Workshop Abstracts (See Appendix)

Pre-workshop survey (See Appendix)

We sent participants a survey to assess their digital needs in research data, teaching, and publishing. The survey targeted the three main areas of our focus and sought to provide us with an overview of main trends and needs among the participant pool. An anonymized result of the survey is available upon request. Thirty eight participants completed the survey and, while the survey results were helpful for framing the discussions at the workshop, in order to make more generalized assertions we need the participation of a larger group.

Linked Open Data Overview (See Appendix)

The short overview of LOD provided outlines the basics, the infrastructure needs, as well as the potential of LOD for librarians and medievalists, with specific examples drawn mostly from the field of medieval studies.[3]

Workshop Organization

[3] See also W3C's "What is Linked Data?" in http://www.w3.org/standards/semanticweb/data; T. Heath and C. Bizer, *Linked Data: Evolving the Web into a Global Data Space*, Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1 (2011). Available online at http://linkeddatabook.com/editions/1.0/. Another great resource explaining LOD and the processes involved is Seth van Hooland, Ruben Verborgh, *Linked Data for Libraries, Archives and Museums* (London, 2014). Also see the authors' blog: http://book.freeyourmetadata.org/.

The workshop revolved around presentations on Day 1 (May 11, 2015) and breakout sessions on Day 2 (May 12, 2015). The abstracts of the presentations from Day 1 are available in the appendix. During a working lunch on Day 2, current LOD project leads and library information architect teams deliberated the general costs (both intellectual and financial) of hosting and exposing LOD and the technical infrastructure needed to support it. On Day 2, three separate breakout groups discussed specific questions regarding research-related, teaching-related, and publications-related data. On the afternoon of Day 2, each group presented a summary of their deliberations, followed by a wrap up session and an overall assessment of next steps.

SUMMARY OF DAY 1 DISCUSSIONS

The first day's discussion covered a wide array of current trends in digital scholarship and challenges around creating, sharing and preserving digital data in research, teaching, and publication. A 15-minute discussion led by assigned group leaders followed each presentation.

- Geraldine Heng welcomed participants and Ece Turnator presented an overview of LOD and its possible applications in medieval studies, with specific examples drawn from LOD projects relevant to the field. The ability to specify and link different datasets without having to agree upon an ontology or vocabulary emerged as a strength of LOD. The importance of collaboration between the medievalist community as domain experts, librarians, and technologists was underscored. Turnator urged participants to think about resources that may be ripe to be encoded as LOD.

- Maria Esteva's presentation emphasized the importance of metadata in forming LOD resources ("first rich metadata, then LOD"). Current data management trends propose that researchers take over data curation during the research project's evolution, and in many cases beyond, maintain digital scholarly projects online through their institutes or departments. Since most medievalists have no formal training in metadata design, teaming up with librarians and information technology professionals early on ensures that solid standards-compliant metadata practices and adequate technical infrastructure are developed from the beginning of the project. Because metadata creation is complex and requires many iterations and collaborations, Esteva emphasized a modular approach in which metadata creation and use are implemented as

workflows across the research process, both to organize the data during the data gathering and study phases and to represent it at the online publication stage. Those workflows then map to and function within infrastructure resources. Such a platform will necessarily involve some degree of automation for basic metadata gathering, to address data scalability as well as to ease the burden on the individual researcher. Interpretative data description—the strength of the researcher—may be added towards the end of the project prior to publication.

- Leif Isaksen's presentation (available online) offered an overview of an active LOD project that is relevant to medieval studies, the Pelagios Project.[4] Pelagios links data from over 39 partner LOD projects and their annotations using place/location as the common denominator between them. The presentation emphasized the importance of annotation and preserving those annotations in LOD format, rather than attempting to create a single unified model via standard vocabularies and ontologies that every scholar would be expected to abide by. The outcome is a promising and impressive approach for medieval scholarship that not only disambiguates and links together data, but also creates a platform for scholars to rapidly discover data that they initially did not know existed.

  Isaksen highlighted that, when researchers choose to participate in projects like Pelagios' Recogito, they have the opportunity to immediately start contributing to an LOD site without having to build it. In doing so, they join the team of collaborators who currently utilize existing LOD sites and stable web addresses in research. The Recogito platform for annotating historical texts and images gives scholars working in the Latin, Chinese, Greek, Arabic, and Persian traditions a means to disambiguate and specify location names via LOD annotations.

- Geraldine Heng offered an overview of the history of the Global Middle Ages Project. Ece Turnator discussed the focus of the Global Middle Ages portal and its potential for linking various medieval projects that aim to create a better understanding of the medieval world. The remainder of the presentation

[4] "What Do You Do With a Million Links?"
https://dl.dropboxusercontent.com/u/18447022/Pelagios_WDYDWAML_SCS.pptx

focused on the workshop survey results, highlighting trends in usage of digital resources by medievalists.

- Eric Kansa introduced Open Context, another case study of an LOD project centered on publishing anthropological and archaeological data on the web. Kansa highlighted the limitations of a Taylorist approach in attempting to serialize, centralize, and measure research data.[5] He then stressed that the LOD approach requires a new and different mindset than the one in which the research community has been trained so far—one that considers the data creation process an endless chain of iterations and places the researcher in an ecosystem of communities built on mutual trust and a willingness to share data.

- Adam Rabinowitz described Periods, Organized (PeriodO), an NEH-funded LOD project which aims to build a gazetteer of period definitions—specifically, the association of a given period term with explicit coordinates in time and space) produced by authorities. Each definition will have its own stable web address (URI) so that scholars will specify which definition of a period they are referring to by copying and pasting the PeriodO URI into their work. This simple act removes the ambiguity in the period claims scholars make in their own work by specifying which period scheme they are using in a given context.

- Lilla Kopár, presented the NEH-funded digital portal, Project Andvari, developed in cooperation with IATH/UVA (The Institute for Advanced Technology in the Humanities, University of Virginia) and multiple collaborators, such as the British Museum, Kringla (the portal for the Swedish National Heritage Board)[6], and the Norwich Castle Museum. This project provides integrated access to early medieval art and artifacts from northern Europe. Project Andvari is an impressive example of collaboration among museums, libraries, and scholars where museum and library materials benefit from Andvari's enhanced metadata and sophisticated faceted search options, while scholars benefit from having a single portal to access resources that reside in various repositories. An additional benefit of Andvari is its prompted semantic search function that, in a rather serendipitous manner, allows researchers to stumble upon semantically related data or objects that they were not necessarily searching for. In addition, the project collaborates with

[5] On Taylorism see: http://en.wikipedia.org/wiki/Scientific_management.

[6] http://www.kringla.nu/kringla/.

The British Museum's crowdsourcing tool MicroPasts, in order to test its newly developed thesaurus for iconography and to tag relevant sample objects for Project Andvari.

COMMON DISCUSSION POINTS

What is LOD?; how do we produce it and how do we use it?

On a higher level, LOD is about sensibly curating the (over)abundance of data so users can find the semantically same needle they were looking for among the hundreds of possible others that are spelled similarly, sound alike, or look alike. Creating LOD involves assigning a single, stable web identifier for a single resource, no matter what we humans call it. For example, if researchers use a multiplicity of names for wild sheep, the array of terms can be reconciled as one web address, as seen in the Encyclopedia of Life.

As a more medieval studies-related example, there are many cities named Antioch in the ancient/medieval world. On a simple level, if a scholar would like to specify which Antioch they are using in their work, they could look it up on Pleiades—an LOD gazetteer which curates unique web addresses for geographical places—and copy-and-paste the web address into their work.[7]

The large gazetteer model (reference-centric) and the subject-specific (property-centric) model.[8]

There are at least two well-known models of the kinds of data that can be exposed as LOD. Both models were represented at the workshop. The Stanford Linked Data

[7] A gazetteer, in "Pleiades-sense," is a directory that assigns URIs to place concepts. A gazetteer is a directory, much like a phone directory; it contains reference information. For a recent application of LOD in gazetteer format on a biodiversity collection see, Silvio D. Cardoso, Flor K. Amanqui, Kleberson J.A. Serique, José L.C. dos Santos, Dilvan A. Moreira, "SWI: A Semantic Web Interactive Gazetteer to support Linked Open Data," in *Future Generation Computer Systems* (accepted May 6, 2015), 1-11. DOI: http://dx.doi.org/10.1016/j.future.2015.

[8] Leif Isaksen's definitions are placed in parentheses. According to Isaksen, the reference-centric model, like Pelagios, tends to relate materials based on references like people, places and things, often with very simple (or even semantically opaque) relationships. The latter, i.e. the property-centric model, emphasizes the nature of relationships between entities, such as "producedBy," "mintedAt," etc. We would like to underscore that these models are not mutually exclusive; a project may attempt to achieve both. Generally speaking, reference-centrism supports improved resource discovery and is easier to scale across heterogeneous data. Property-centrism is better suited to inferencing/complex queries and often easier to derive from individual data sets.

Workshop of 2011[9] describes the alignment of different metadata models into an interoperable collection of Linked Data as a "major challenge." Since then, at least in the fields of classics and medieval studies, some models have emerged that successfully tackle this challenge. For example, the gazetteer model, aims to answer the "big bucket" questions of "who, what, where, and when," by providing a resource similar to a phone book or a lookup table for these big questions that are essential for many projects, irrespective of field.[10] The Pleiades and PeriodO projects provide gazetteers for an expanding number of geographical places associated with the ancient world and an unlimited number of historical period claims made by the scholarly community about any part of the world, respectively. The Pelagios Project's partners offer a range of sites that either currently produce LOD relating to medieval studies, or are examples of possible future applications of LOD to medieval studies.[11] Among these partners are projects that constitute examples of the subject-specific model, such as Nomisma and Epigraphic Database Heidelberg. These resources document mint and inscription information tied to a specific geographic location in the ancient world, respectively. The Pelagios Project, on the other hand, aggregates the results from partner LOD sites. For example, if researchers look up a specific place under the Pleiades API,[12] they will have access to the links from partner sites that relate content to that particular geographic location.

One such example of the non-gazetteer/subject-specific model represented at the conference is Open Context, a publishing platform/data repository for anthropological and archaeological data. Open Context provides access to raw data and makes it possible to relate this data to other research data. Open Context, in return for a one-time fee, cleans up data using Open Refine and serializes the data, a process that entails putting the data into the same metadata format that all other data in Open Context uses, enabling on-demand data harvesting.[13] Open Context provides a faceted

[9] Report of the Stanford Linked Data Workshop, 27 June-1 July 2011, 37.

[10] For other examples please see: http://www.geohumanities.org/inaugural-meeting.

[11] A great resource replete with relevant case studies is NYU Ancient World Digital Library's 2014 co-publication with the Institute for the Study of the Ancient World, "Current Practice in Linked Open Data for the Ancient World," http://dlib.nyu.edu/awdl/isaw/isaw-papers/7.

[12] API stands for Application Programming Interface: https://en.wikipedia.org/wiki/Application_programming_interface.

[13] Open Context's process from acquisition of a dataset: 1. Load data into Open Refine and clean it up there; 2. Use Open Refine API and load it into Python, programming language; 3. Metadata processing, putting the cleaned up new dataset into Open Context's metadata schemata.

search option and also makes data available on Github for those who have the programming skills to apply data analysis techniques and ask specific research questions of the data.[14]

What skills are needed? Do medievalists have those skills?

Creating LOD requires a creative mix of domain-specific knowledge, programming expertise, and institutional commitment, leading to a more complex and collaborative type of work. The end result, after deliberation among domain, content, and technology experts, is putting web identifiers into machine-readable (i.e. serialized) data files that define the resource and link it to other web resources, typically using the RDF metadata model.

This means that, when other institutions also serve their resources in computer-readable files with unique identifiers, these identifiers can be linked, thus enabling disambiguation and discovery.

In some cases, the serialized data lives in a JSON file which can be manipulated with most programming languages (Javascript, PHP, Python, Ruby, etc.) immediately. Other projects use an older model, XML, for traditional serialization format. Some LOD projects use JSON-LD, a version of JSON that can also be expressed as RDF. This direct manipulation of the data requires that researchers learn how to program, depending on the tools provided to researchers via an API—as in the Pelagios and Open Context projects, which can ease access to data—creating a significant challenge impending further diffusion of LOD projects: either the search and find options must be very easy and applicable for non-programmers and their research questions, or scholars need to learn programming. In short, LOD poses a significant technical challenge for scholars, who often graduate without learning a programming language.

Another important aspect of LOD is that it is not a tool that can be used "out of the box": it is a resource that needs to be collectively and constantly created, used, and linked both to and from. It is a blessing and a curse of LOD that its benefits are mainly theoretical until large amounts of linked data become available, which in turn

[14] Ben Marwick's project to integrate Open Context's API into an RStats package: https://github.com/ropensci/opencontext.

requires a research community trained in the specific skill sets necessary to use linked data.[15]

For LOD to succeed, the benefits and challenges of producing and using LOD resources need to be communicated to all stakeholders in the research community[16] and an incremental approach needs to be adopted to provide genuine benefits within realistic timescales. Like the Web of Documents, the Semantic Web needs to grow organically, depending on needs and opportunities as they arise.[17]

LOD requires a mindset change, a recalibration of our old value system:

Rather than the development of particular tools for particular systems, LOD is fundamentally about collaboration and thrives in an ecosystem where research teams see themselves as stewards and *curators*, not *owners*, of their research data. It functions best when researchers are open about research outcomes so that others may use and build upon their work.[18] In this sense, LOD foregrounds the interests of the community of practice, rather than the individual researcher.

However, the change of mindset and practice required by LOD should not be seen as asking researchers to "be a good person and reap no reward." At its best, LOD

---

[15] The library community has started working, so far successfully, in "creating identifiers for entities that do not yet have them" for the purpose of enlarging the pool of data to disambiguate them later. See, Karen Smith-Yoshimura, "Shift to Linked Data for Production," in http://hangingtogether.org/?p=5195 (May 13, 2015). Also see, Dean Kraft and Simeon Warner, "Linked Data for Libraries (LD4L)" in http://www.dlib.org/dlib/may15/05inbrief.html (May/June 2015). We would like to thank Adam Brin for both references. On June 4, 2015 Coalition for Networked Information (CNI) published "Bibflow: A Roadmap for Library Linked Data implementation," is available at https://youtu.be/Z-g_yJ1FL0U and https://vimeo.com/129797055.

[16] See Peter Murray's blog on "Institution-wide ORCID Adoption Test in U.K. Shows Promise" (May 22, 2015): http://dltj.org/article/institution-wide-orcid-adoption-test-u-k-shows-promise/.We thank Adam Brin for the reference.

[17] On the transition from the Web of Documents to Linked Data see, G. Képéklian, O. Curé, L. Bihanic, "From the Web of Documents to the Linked Data," in E. Zimányi, R.-D. Kutshe (eds.), *Business Intelligence. 4th European Summer School, eBISS 2014, Berlin, Germany, July 6-11, 2014 Tutorial Lectures* (Brussels, Berlin, 2015), 60-87. http://link.springer.com/chapter/10.1007%2F978-3-319-17551-5_3#page-2. DOI: 10.1007/978-3-319-17551-5_3.

[18] A more nuanced discussion on the necessity of planning the end of digital projects from the start and the possible roles scholars, librarians, technologists could play in that ecosystem see, Miriam Posner, "Data Preservation and Stewardship" in http://miriamposner.com/blog/digital-humanities-and-the-library/#datapres.

increases the impact of scholars' work exponentially, which, if measured and framed correctly, could help scholars' careers in concrete ways.[19]

Is LOD worth the intellectual and financial investments, despite the challenges?

Yes, because the web and the digital potentials it offers are here to stay. Web 2.0, the Semantic Web, is a relatively new reality with immense potential even though conceptually it dates back to the 1960s.[20] The example projects outlined above constitute strong evidence that LOD is valuable because it enables users to disambiguate and link stable online resources to enable serendipitous discovery through the linking. The essence of academic discourse is connectivity, whether through citations, footnotes, cataloguing systems, indices, or other mechanisms. LOD provides an additional means of doing so. Indeed, it is hard to think of any academic work existing in isolation. In this context, LOD is simply an attempt to support these principles within the contemporary medium of the Web.[21]

LOD and ambiguity; LOD and small data; LOD and the "edge cases":

Questions arise in scholarly minds: can LOD work with small, fragmentary, ambiguous, or "fuzzy" (i.e., this fragment is from "the north of France?") data sets? Does the scope and size of materials matter? How does LOD help "fuzzy" and small data sets, and how can those of us working with—and creating—those datasets efficiently integrate LOD into our research and our assignments to keep building up the LOD world?

There is no simple solution. Standards are interpretive: their application depends on each project and each research group, and what works for small datasets will likely not scale to bigger, shared ones. Cross-referencing things or entities (people, events, places, classification schemes) is easier than cross-referencing across properties (i.e. column headings across different tables) which may not be applicable to all users of the data. Linked data is useful for certain kinds of applications, but it is not a magical tool for all of a researcher's interests. The community or communities of scholars

[19] For further discussion, see section on Day 2 below.

[20] http://en.wikipedia.org/wiki/Semantic_Web.

[21] We thank Leif Isaksen his input in the last two sentences.

need to lead the collaborative discussion and collectively decide, ideally at the end of a transparent process, on entities and how to put them into LOD format.[22]

That said, with linked data one can have as many annotations as one wishes. If there are multiple interpretations, it is possible to link to all of them without having to decide upon a specific set (which is essentially what PeriodO is doing by aggregating all scholarly period claims). Still, how we visualize areas of uncertainty must be worked out collectively and openly, informing the community of practice about the decision process behind the choices made in terms of metadata, serialization of data, and programming. It is imperative that the community understands how LOD works. This is particularly true in the case of smaller datasets and serializations of ambiguous or "fuzzy" data; researchers need to know what kinds of decisions are made in the process of the serializing data in LOD format.

The great metadata problem: how granular shall our metadata be?

The metadata experts present at the workshop and, in particular, one of the presenters on Day 1, Maria Esteva, underscored the essential role that metadata plays in scholarship sharing and reuse, and specifically in the serialization process into LOD. Esteva described metadata creation as a process similar to that of research; just as the research process is never linear, so is the metadata process always in flux. Within this context, the recommendation is to have basic metadata at the outset and enrich the data with richer metadata depending on the evolving research needs.

In a distributed curation approach to metadata, conceived by Esteva, stages extend from:

a. technical/preservation metadata which can be programmatically extracted and stored as data is being gathered;
b. basic descriptive (non-granular) metadata, recommended by librarians working with the researchers to organize the data as it is being gathered and shared amongst team members;
c. provenance metadata during the data analysis/study process;

[22] See n. 8 above.

d. “descriptive basic” stage that involves basic (i.e. “non-granular”) metadata for purposes of organizing the data in big buckets of themes during research and for sharing with team members;[23]
e. the “descriptive scholarly” stage which are the sophisticated descriptions of the data that the researchers create based on the results of their studies.

Across all these stages, involvement of a metadata librarian to design workflows and avoid heavy duty curation work at advanced stages of the research process is key.

This collaborative process among medievalists and librarians implies a change in the classic humanities research model in which researchers are typically alone throughout the research process and the end result does not see daylight until publication when, from the metadata perspective, it is often too late for the research data to be formatted so that is usable for other scholars. Granted, this is not true for all kinds of research in medieval studies, however, working within open data parameters[24] requires a radical change from the current humanities model.

Recap: How can we effectively establish an “open and decentralized”[25] Medieval Studies?

- **Clearly communicate of the benefits of LOD.:** Clearly explaining the value of LOD is essential for getting the scholarly community to support it. JISC’s (Joint Information Systems Committee) recent “ORCID Implementation Report” emphasizes the significance of demonstrating by example to faculty and the value in creating and joining research communities that are built on LOD principles.[26]

[23] By less granular, basic metadata organization, Dublin Core [Basic] Metadata Element Set is meant. Furthermore, descriptive basic metadata can be implemented as a recordkeeping system from which tags/labels can be automatically extracted to populate minimum elements in Dublin Core records.

[24] See http://5stardata.info/.

[25] L. Isaksen, *Archaeology and the Semantic Web*. Ph.D. Dissertation, University of Southampton, School of Electronics and Computer Science, 2011 (available at http://eprints.soton.ac.uk/206421/).

[26] The report states that: “Many of the pilot projects found it difficult to articulate the benefits of ORCID to individual academics and researchers. However, most pilots reported that senior management quickly understood the benefits to the institution”; “Some institutions reported having specific

- **Form a community of trust willing to share data, ideas, and research outcomes openly.** This is the *sine qua non* of LOD; therefore, the participants recommended working on LOD projects in small groups with people who are genuinely enthusiastic about collaboration.[27]

- Determine **which entities** are ripe for encoding in LOD as a scholarly community and further develop that community of practice based on mutual trust. This requires open discussion and transparency regarding the decision-making process involved in encoding LOD entities. If this proves to be impossible to achieve, it may be sufficient for small groups to take responsibility for particular resources, such as creating a gazetteer.

- Simplify and demystify the metadata design process. Streamline and delineate the metadata preparation process for the medievalist community.

- Simplify, demystify, and clearly communicate the creation and application processes of LOD research outcomes. Create APIs and provide extensive documentation that will keep the data serialization process transparent. This will translate to easily-understood research outcomes.

SUMMARY OF DAY 2 DISCUSSIONS

Research-Related-Data Group:

Group Lead: Michael Widner (Stanford University)

Christopher Atwood (Indiana University Bloomington)
Alexandra Bolintineanu (University of Toronto)
Adam Brin (The Digital Archaeological Record)
Matthew Evan Davis (North Carolina State University)

---

difficulty in articulating the benefits of ORCID to particular disciplines or user groups." https://repository.jisc.ac.uk/6025/2/Jisc-ARMA-ORCID_final_report.pdf, 22, 23, respectively.

[27] This is not a simple task either largely because humanities faculty and faculty in general often have no training and no interest in collaboration due to lack of institutional incentives. This article from 1998 is still very relevant: S. J. Bowen, J. Stiles "Experimenting with Models of Faculty Collaboration. Factors that Promote their Success," *New Directions for Institutional Research* 1998.10 (1998), 31-55.

Jee-Hyun Davis (University of Texas at Austin)
Roger Martinez (University of Colorado, Colorado Springs)
Kent Norsworthy (University of Texas at Austin)
Dorothy Porter (University of Pennsylvania)

Medieval data has the potential to be so complex that it is fundamentally impossible to capture every bit of complexity in LOD.[28] The granularity of the data should match the level of detail required by the researcher. Highly granular data can be prohibitive to access and reuse for non-experts, or for experts in related but different sub-fields. Some LOD experts recommend gazetteers built with a focus on first providing access to gateway terms such as who, what, where, and when, and then expanding only as needed.

Following the suggestions of the Workshop's Research Data group, we see this paper **as a draft outline of suggested general norms for metadata for medievalists, and not as a set of definitive rules.**

Linked Open Data for Research Data

LOD is a set of practices for publishing data on the web to make it machine readable by using URIs (uniform resource identifiers, of which URLs are a subset) to name things, and by writing statements about those things in order to explicitly state their relationships to one another. The Leif Isaksen, Adam Rabinowitz, and Eric Kansa presentations, and the discussions that ensued, presented linked data (the semantic web) as a solution to the "too much data" problem and as a way to make amorphous big data intelligible, reusable, and open to computational knowledge discovery.

Linked Open Data and Medieval Studies

As our discussions and presentations on Day 1 showed, LOD projects have flourished in Classics and Archaeology.[29] Unlike Classics, Medieval Studies does not currently

[28] According to Leif Isaksen, "capturing all layers of complexity as LOD is fundamentally impossible because text/data presupposes interpretation (and all LOD is ultimately interpretative). While it is still important to discuss how much complexity is useful in a specific context, this should not be generalized."

[29] LOD projects in the humanities are not limited to these two sub-fields. Stephen Ross, Professor of English at the University of Victoria and the General Editor of the Routledge Encyclopedia of Modernism online, was awarded a grant to create a faceted search interface based on RDF and LOD through the Linked Modernisms Project, for which see, http://src-online.ca/index.php/src/article/view/186/370.

possess comparable projects, though this is fast-changing. Few dedicated digital gazetteers cover the people, places, and periods of the medieval world, but the benefits of LOD are applicable to the medieval studies community and include enhanced discoverability, sharing, and reuse of research data. More importantly, alignment with the library world is valuable because libraries are increasingly moving towards linked data (The Library of Congress, Digital Public Library of America) to make their resources more readily discoverable on the web. Of all these potential benefits, our main need is to make our data discoverable: if we can find other projects in our area, we can build communities of practice, and we can build links among our projects.

Workshop participants noted that LOD is most usefully viewed not as a binary ("is LOD/is not LOD") but as a spectrum: at the lowest level, data that is simply available on the web, ideally, with an open license, provides the benefits of discoverability, sharing, and reuse; at the highest level, data that is available under an open license, encoded to open W3C standards (RDF and SPARQL), and linked to other data, provides the most powerful benefits, but also requires the greatest investment in terms of labor and technical expertise. The discussion below is structured along these levels and aims to provide best practices for achieving each level of openness.[30]

Data: Content

As medievalists, our research data is the material of our studies: historical and fictional persons and places; literary and historical texts, written in manuscripts, on church walls, or carved into standing crosses; material artifacts, from coins to cathedrals; archaeological sites; and the philosophical, religious, and cultural concepts, dynamics, and narratives we interpret. Of these, the interpretive categories are hardest to encode as linked open data; no ontology exists that can capture multi-vocal, theoretical scholarly conversations, debates, disagreements, and narratives.

What we can productively encode, link, and render machine-discoverable via the semantic web are the simplest, most granular data points: people, places, things, and times. We need to track this data and encode it as LOD by drawing on name

Also see, New York Public Library's Linked Jazz Project: https://linkedjazz.org/about-the-project/ to name a few other projects among a growing interest within the humanities community overall.

[30] See http://5stardata.info/. Also see, "Linked Open Data Overview" in the Appendix.

authorities and gazetteers to unambiguously identify our data and to assign them URIs.

Data: Format

Both the pre-workshop survey as well as an informal survey of medievalist workshop participants revealed that our digital work proceeds in a diversity of formats and software platforms, depending on the nature of research and researcher preferences. To manage this wilderness of difference, we recommend the pairing of **durable structured data with a lightweight interface.**

With this pairing, we should aim to make the data extractable in the simplest medium possible (ideally, structured: CSV, TEI P5, RDF, etc.), so that the data can live independent of any platform and can thus also be ingested into any other platform.

Metadata

To ensure that we can make our data discoverable, share and preserve them, and keep them intelligible, we need to collect data about our data and generate basic descriptive metadata (focusing on such fundamental categories as persons, objects, time periods, and places); metadata about intellectual property, rights, and licensed usage; and technical preservation metadata.

Basic descriptive metadata can be expressed in Dublin Core, a metadata schema created to capture "core metadata" for simple and generic resource descriptions. Dublin Core has been ratified in accordance with major international standards (IETF RFC 5013, ANSI/NISO Standard Z39.85-2007, and ISO Standard 15836:2009) and is widely used in the library community.[31]

Metadata about intellectual property, rights, and licensed usage also defines the terms on which we share our data and allow their reuse in the wider community. Creative Commons offers widely accepted copyrights licensing standards for sharing resources on the web. [32]

[31] In order to do this in LOD one could use the Dublin Core terms ontology for the properties (http://dublincore.org/documents/dcmi-terms/) and, where available, use URIs as values. We would like to thank Leif Isaksen for this comment.

[32] https://creativecommons.org/about.

Finally, preservation metadata collects technical and administrative information required to keep the data usable and accessible. Specifically, it "stores technical details on the format, structure and use of the digital content, the history of all actions performed on the resource including changes and decisions, the authenticity information such as technical features or custody history, and the responsibilities and rights information applicable to preservation actions."[33]
Preservation metadata is usually supplied by librarians. The PREMIS Data Dictionary for Preservation Metadata is a relevant internationally accepted standard.[34]

The preparation of research data for LOD encoding requires the expertise both of the researchers who collect and use the data, and of librarians who curate the data.[35]

As Maria Esteva noted in her presentation, high-quality metadata is a process that takes place throughout the research cycle, in collaborations between researchers and librarians. We recommend that researchers and librarians follow Esteva's suggested stages below in their collaboration:

1. Identify the stages of the research; describe what kinds of data will be collected, created, or needed at each stage.
2. Identify the researchers' data publication and reuse goals: is the data intended for sharing and publication? If so, the data meets the basic requirement to be encoded as linked data.
3. Define metadata standards and modules: what kinds of metadata will researchers and librarians create and archive? What standards can best serve the project's purposes? For example, a standard like Dublin Core is more basic and less granular; a standard like MODS is more granular, and accordingly more labor-intensive.[36] What purposes does each module of metadata share?

---

[33] http://en.wikipedia.org/wiki/Preservation_metadata#cite_note-PADI-1.

[34] http://www.loc.gov/standards/premis/v2/premis-2-0.pdf.

[35] On the complexities of data curation throughout the research cycle, especially for highly interpretive humanities' data, see, for example, Trevor Muñoz and Allen Renear, "Issues in Digital Humanities Data Curation," (white paper, 2011): http://hdl.handle.net/2142/30852); Tobias Blanke, Mark Hedges, and Stuart Dunn, "Arts and humanities e-science—Current practices and future challenges," *Future Generation Computer Systems* 25, no. 4 (2009): 474-80; C. L. Palmer, N. M. Weber, T. Muñoz and A. H. Renear, "Foundations of data curation: The pedagogy and practice of 'purposeful work' with research data" *Archive Journal* 3 (2013): http://www.archivejournal.net/issue/3/archives-remixed/foundations-of-data-curation-the-pedagogy-and-practice-of-purposeful-work-with-research-data/.

[36] http://www.loc.gov/standards/mods/.

4. Identify the technical infrastructure: what are the tools and platforms in which researchers will create, store, process, and archive this metadata?
5. Map the data to a collection architecture: decide where the data will live at different points of the research lifecycle.
    5.1. As soon as data is collected or generated, it should be roughly categorized ("big buckets"). Labelled with basic metadata, it can go both into the archive and back to researchers for elaboration.
    5.2. Researchers take this pre-organized data and add more sophisticated and interpretive metadata based on their research. Labelled with this metadata, the data can go both into the archive and back to researchers.
    5.3. At each of these stages, data is archived.
    5.4. At each of these stages, data can both go into wider circulation (exposed to publication, visualization, analysis, sharing) and into the archive.
6. Realize that metadata is not an end product: metadata is a process, developed throughout the research cycle, through the collaborations of researchers and librarians.
    6.1. Researchers act as data generators and curators, with their domain expertise informing metadata creation from the outset.
    6.2. Librarians and archivists provide synthesis, structure, and archiving on the fly, ensuring data quality and preservation throughout the research cycle.

Encouraging Data Reuse

Basic LOD, as described above, serves a primary goal for researchers—rendering data more readily discoverable. Additionally, in order to create strong communities of practice, data reuse needs to be encouraged. To do this, we must surmount technical and social barriers.

Technical Barriers

This hurdle is two-fold: data sets may not be in a reusable format, or the expense (in terms of labor and technical expertise required) to render the data technically reusable may be too high. To solve this problem, we propose using Dublin Core as a lowest common denominator. For some smaller projects Omeka could be used as a

data collection platform.[37] Omeka is a widely accepted, user-friendly tool for creating online exhibitions, and it has traction throughout the humanities.[38]

Social Barriers

More formidable are the social obstacles to data reuse: **there are no professional mechanisms in our discipline yet for appropriately valuing, peer-reviewing, and rewarding the considerable scholarly labor of making high-quality research data available for sharing and reuse**. Workshop participants noted that in archaeology, the work of collecting and cataloguing archaeological findings, and of making this data available in robust digital form, is not seen as intellectual work of the same value as more traditional scholarship and, as a result, it does not contribute to professional advancement. In medieval studies, similarly, the philological scholarship of the late nineteenth and early twentieth century laid the groundwork for the interpretive work of the late twentieth and early twenty-first centuries. However, during the same period, philology itself lost prestige. Digital medieval studies is in an analogous position. Although it remains necessary to build the foundations (digitized and transcribed manuscripts, digital scholarly editions, etc.) and technical infrastructure for future work, the academy in general does not value such work as prestigious or important. For faculty inclined to do such work to still receive tenure and promotion, they must "do twice the work" by also publishing traditional print scholarship about their digital labors.[39]

---

[37] We would like to underscore that very few (if any) data standards guarantee re-use without significant reprocessing. The problem is that humans seek patterns in the phenomena represented by data, not the data itself. For example, the typo in string "1 2 3 4 S 6 7 8 9" is obvious to a human but not to a computer. Providing machine-readable data is putting data in a format that computers can process. Ultimately, we need to come to the realization that carrying out sophisticated analyses of digital data will likely always require a certain amount of processing and manual labor. We thank Leif Isaksen for this valuable insight.

[38] Omeka also has technical affordances that support collaboration and, even more importantly, acknowledgement of each collaborator's labor; it enables the programmatic extraction of data via its API (Application Programming Interface: https://en.wikipedia.org/wiki/Application_programming_interface) and the exposure of its metadata via OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting: https://www.openarchives.org/OAI/OAI-organization.php); in its current version (Omeka 2.x), it supports the export of data as JSON and RDF-XML; and in its soon-to-be-released version, Omeka-S, it will support JSON-LD data.

[39] Sydni Dunn, "Digital Humanists: If You Want Tenure, Do Twice the Work" *Chronicle Vitae,* Jan. 14, 2014. https://chroniclevitae.com/news/249-digital-humanists-if-you-want-tenure-do-double-the-work

To encourage data reuse, we have the following suggestions:

1. Create a portfolio of "success stories" with qualitative and quantitative evidence demonstrating the benefits of data reuse (qualitative evidence for scholars and quantitative evidence for administrators).[40]
2. Acknowledge the ethical ramifications around data sharing. Since there are currently no mechanisms in Medieval Studies for acknowledging and rewarding the labor that goes into collecting high-quality data and making that data available, scholars must be supported both professionally and institutionally in making informed choices as to whether to share or not to share their data.[41]
3. Persuade major disciplinary organizations (the MLA and the Medieval Academy) of the value of data publication and data reuse, and of the importance of creating reward mechanisms that acknowledge this value.[42]
4. Reach out to the scholarly community through conference presentations (specifically, digital project presentations that focus on the consequential scholarly questions rather than the digital methodology), white papers demystifying linked open data for scholarly use, and pedagogical data reuse. For the latter, see the story of PeriodO as a digital data publication project arising directly from pedagogical questions and project APRICOT, a digital hub for collecting and reusing teaching resources.[43]
5. Institutionally, too, social obstacles to overcome include a lack of funding for data collecting and digitizing of medieval collections and a lack of

---

[40] JISC's "Institutional ORCID Implementation and Cost-Benefit Analysis Report", cited in n. 26 above, notes the contrast between the difficulty of communicating the benefits of ORCID to individual academics and the ease of communicating the benefits to senior management. Different communication strategies—some focusing on qualitative, theoretical, discipline-driven analyses and research success stories, others on quantitative benefits to institutions—are needed to address different audiences. One such effort in the field of archaeology is CAA's (Computer Applications and Quantitative Methods in Archaeology) Recycle Award which "seeks to recognize those who breathe new life into old data." We thank Adam Rabinowitz for this reference.

[41] Christine L. Borgman, "The Conundrum of Sharing Research Data." *Journal of the American Society for Information Science and Technology* 63, no. 6 (2012): 1059-1078; Carol Tenopir, Suzie Allard, Kimberly L. Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame, "Data sharing by scientists: practices and perceptions," *PLoS One* 6, no. 6 (2011).

[42] See "Guidelines for Evaluating Work in Digital Humanities and Digital Media", https://www.mla.org/guidelines_evaluation_digital, especially p. 2; B. Nowviskie, "#alt-ac: Alternative Academic Careers for Humanities Scholars." http://nowviskie.org/2010/alt-ac/ (January, 2010).

[43] T. Rose-Steel, et. al. "Forging a Pedagogical Hub for Medieval and Early Modern Teaching. APRICOT, A Peer-reviewed Interdisciplinary Collection of Objects for Teaching" (forthcoming).

communication between research faculty and the library.[44] To work around these obstacles, we recommend organizing both formal and informal meetings between librarians and researchers who have or wish to create digital data—meetings where each party's expertise is recognized and where unproductive institutional hierarchies are not reinstated. These in-person meetings will allow participants to learn about one another's work, goals, and work culture, and lay the groundwork for collaboration throughout the research process.

Teaching-Related-Data Group:

Group Lead: Tamsyn Rose-Steel (Johns Hopkins University)

Melanie Cofield (University of Texas at Austin)
Monica Green (Arizona State University)
Jennifer Hecker (University of Texas at Austin)
Thomas Kealy (Colby Sawyer College)
Lee Mordechai (Princeton University)
Tim Pauketat (University of Illinois at Urbana-Champaign)
Adam Rabinowitz (University of Texas at Austin)
Lynn Ramey (Vanderbilt University)
Bridget Whearty (Stanford University)

In "A Letter to the Humanities: DH Will Not Save You," Adeline Koh argues that Digital Humanities must focus more on what the humanities already do well, in particular, critically inspecting "identity categories" and "access and privilege in the process of making." The process of critical making in the humanities, Koh insists, is only humanistic if we also pay attention not just to the methodology but to the *consequences* of its methodology, its "social, political and economic underpinnings." Koh singles out "digital work that focuses more on culture than computation," "digital recovery efforts of works by people of color," and "work with undergraduates and digital pedagogy" as particularly important—and sadly neglected—areas of growth.[45]

[44] See n. 35 above.

[45] http://www.hybridpedagogy.com/journal/a-letter-to-the-humanities-dh-will-not-save-you/.

Koh's article provided the shared foundation for our discussion on pedagogy, LOD, and the different roles that LOD can play in our work as instructors. We focused on three different roles:

1. Discoverability. LOD as an aid for instructor's discovery and use of online resources, as well as an aid for students' discovery and for fostering information literacy.
2. Experiential learning. LOD as an opportunity for "critical making" assignments, in which students don't simply use LOD made by scholars but participate in the making and maintenance of LOD tools and resources.
3. Community. LOD as an opportunity for scholars and students to work collaboratively as co-creators of knowledge with librarians and archivists, in which scholarship and pedagogy are designed to foster the tools, data sets, and practices that our library colleagues are in the process of creating.

What became clear in our discussion is that LOD is not a magic wand for creating great pedagogy, nor is it *necessary* for generating good pedagogy. We do not have to bring LOD-focused activities (teaching the importance of the Semantic Web as a publishing platform and the role of URIs in it, as well as serializing the data) into our classrooms to achieve solid pedagogical aims. However, it may be important to highlight the significance of discovery, linking, and reuse as we work with students. In this strict sense LOD may be used as a tool for experiential learning—for getting students to realize the importance of the Semantic Web and how to make the best use of it by teaching them the real-world impact their school work can have when they understand how the Web operates. Responsibility became *the* watchword and refrain of our discussion. It came to stand for two separate but intimately linked principles: first, that any use of LOD in undergraduate curriculum has to have a sound pedagogical rationale behind it and, second, that LOD can be productively used to give students a sense of responsibility for their learning, for knowledge production, and for their field or discipline.

Linked Open Data and Discovery

As researchers and instructors we have a duty to provide "good" data to our communities. We also have a deep need for access to the quality data, quality assignments, and quality resources that our colleagues have created and shared. LOD can assist with this by providing information that has a meaningful connection to other sets of information. For example, a teacher looking for innovative assignments on Chaucer's *Canterbury Tales* might turn to Google or another search engine and

have to dig through the masses of key-word identified information (good and bad) that follows.[46] This teacher might also turn to exclusively-physical teaching resources like *Studies in Medieval and Renaissance Teaching* or sources like *The Rough Magic* which exist on the web, but are not linked out for efficient discoverability. One way that LOD could revolutionize pedagogical practices would be to help connect these valuable and often isolated resources to make them discoverable in a larger and more impactful way.

While instructors might benefit from using LOD to connect the quality resources that are too often scattered and hard-to-find, students have rather different needs. First, and this goes much farther than our discussions on LOD, students need access to quality, vetted facts. College students, particularly those in the United States who have come of age as thinkers in public schools profoundly shaped by the *No Child Left Behind Act* (2001), arrive in college and university settings trained to believe that "learning" is finding an answer and repeating that answer. LOD can play a role in ensuring that the facts our students turn to in their first forays on a subject are substantive, authoritative, and backed by experts in the field. Ideally, searches carried out by students on the internet or library catalogue will lead them to a significantly more useful and relevant web of interconnected information (with those connections themselves carefully curated), rather than a list of keyword results that could vary widely in their authority as well as in their applicability to the subject at hand. While this will not automate information literacy (our group agreed, vigorously, that we will always need librarians and instructors to help students learn to find, sort, and train to critically sort through the overabundance of available information) it does go some way to helping students assess "good" data—from which instructors can help them grow vitally important critical evaluation skills.[47]

Linked Open Data and Experiential Learning

[46] The Canterbury Tales has been the subject of several efforts of various kinds, some specifically for teaching. To name a few seehttp://www.unc.edu/depts/chaucer/chpages.htm and http://www.unc.edu/depts/chaucer/chteach.htm. We thank Lilla Kopár for these references.

[47] That being said, none of this, including LOD, should be seen as an antidote to student access to non-quality facts that exist on the Web, however. We need to acknowledge that there are limits to how well data quality can be monitored or evaluated. Improved techniques for managing issues of "trust" may evolve, but for now it is best to train students to approach *all* information critically than attempt to provide "data gardens" in which we deem it safe for them to play. We thank Leif Isaksen for this insight.

One of the most exciting applications of LOD in pedagogy came from students learning by doing—which is to say, training students to intelligently consume *and also create* LOD.

One of the central tenets of *experiential learning* is that students learn best when they are doing more than responding to hypothetical exercises, exams, and essay questions. (Although, connected to solid learning objectives and outcomes, these all continue to be valuable educational tools.) In the STEM fields, and in education, archaeology, and foreign language study, what learning by doing means may be rather obvious. However, it is sometimes less clear in Medieval Studies what our students should actually *be* doing when we set them to learning by doing. Various approaches have proved successful here, from study abroad programs to student-led dramatizations of *The Canterbury Tales* and class field trips dedicated to experiential learning with period technologies.[48] Another approach is to have students learn by doing scholarly labor, having them take on and experience the role of scholar/expert creating real scholarly resources. This is where teaching with LOD can fit in.

One instructor in the group, Adam Rabinowitz, experiments with assignments that have students work in teams to contribute to the Pleiades gazetteer of ancient places.[49] The Pleiades gazetteer, which provides URIs that unambiguously identify ancient places and associate them with geographical locations and various names used in ancient or modern sources, currently serves as a central node in the graph of Linked Ancient World Data. A Pleiades URI attached to a database entry about an object from, say, Athens, allows the creation of a machine-readable semantic connection between that database and another that refers to the same URI. By asking students to contribute new places to the gazetteer, Rabinowitz challenges students to consider questions of evidence and authority, increase their understanding of

---

[48] See, for example, Mickey Sweeney, "Generating Enthusiasm: Performing Chaucer in the Small Liberal Arts College Classroom," *Studies in Medieval and Renaissance Teaching* 15.1 (Spring 2008); Lisa Darien, "Bridging the Gap: Getting Medieval at the Small Liberal Arts College," *Studies in Medieval and Renaissance Teaching* 14.2 (Fall 2007); Ronald Stottlemyer, "A Study-Abroad Course in Anglo-Saxon Culture: On-Site Experiential Learning," *Studies in Medieval and Renaissance Teaching* 14.2 (Fall 2007).

[49] See for example the published results of an assignment in a Greek Archaeology course that asked student project groups to contribute entries on ancient monuments to the Pleiades gazetteer, here summarized in a Google map: https://www.google.com/maps/d/viewer?mid=zJaj4-9GGbZs.kd264smJwr40&usp=sharing. For a more general discussion on using interactive digital tools in the classroom, see:
http://hestia.open.ac.uk/reading-herodotus-spatially-inthe-undergraduate-classroom-part--i/;
http://hestia.open.ac.uk/reading-herodotusspatially-in-the-undergraduate-classroom-part--ii/;
http://hestia.open.ac.uk/readingherodotus-spatially-in-an-undergraduate-classroom-part--iii/.

scholarly research, confront the complexity of concepts of space and place, and learn how the URIs they create fit into the wider scholarly environment of ancient world data on the web—key critical skills required both in and beyond the academy. In this LOD project, students experience some part of the work of being an expert/academic. They learn how disciplines are shaped by their evidence and methodologies, and come to see themselves as part of the fabric of knowledge creation. Students also learn to discern between how authoritative knowledge is created and discovered versus the information glut that the Web houses. Finally, they learn to take their own expertise seriously.

Projects like these use LOD-centric assignments to put students at the center of the research process. By collaborating on contributing to part of a real research tool in their final projects, students' work has a real, measurable impact upon other researchers and students who will use this tool in the future. This expands student responsibility beyond owning their individual learning process to include a sense of ownership of, and impact on, the larger discipline and field of study. This naturally raises the stakes of this kind of assignment—for students and for instructors. One cannot produce bad data for permanent consumption by a community. Gazetteer-contribution assignments (like all kinds of experiential learning assignments) involve a great deal of work on the part of the instructor, but, as various discussants emphasized, given guidance and carefully-crafted intellectual laddering, students tend to react favorably to these heightened stakes and expectations. When challenged in this way, undergraduate students can produce high-quality work that helps grow the expansive data sets on which the long-term success of LOD depends.[50]

Assignments that engage in knowledge creation through instructor-curated student contributions to LOD gazetteers also provide students with a citable publication. Class projects to create websites, blogs, or online exhibits can provide students with a digital object to put on their CVs, but the longer-term curation and preservation of these assignments varies highly from instructor to instructor, and from institution to institution. Taking part in the creation of LOD can give students the chance to contribute to something more lasting, to which their name can be permanently linked. Of course, LOD is not the only way to achieve student research and publications, but it is one way instructors give students opportunities to participate in coursework that is lasting and helps grow the discipline.

[50] Rabinowitz reports that of the 38 assignments, 37 could be published for scholarly consumption with only minor editorial interventions.

More broadly, for any online digital assignment, students and instructors can generate their own LOD researcher IDs (through gazetteers like Orcid[51] and Vivo[52]) to link together their own scholarly productions. If one of the key goals of LOD is to generate and maintain meaningful connections between pieces of information, then having students create and use their own researcher IDs helps them link up their own contributions, transforming their critical making from discrete units into a meaningful web of work.

Experiential learning can and does take many different forms: from having students roll up their sleeves and learn to make houses using medieval building techniques, to physically or digitally mapping literary itineraries and creating a new LOD identifier for an important monument. LOD is not a panacea; it is, however, a profoundly powerful tool for one of *the* major objectives of a college education: helping students cross the gap between "other people make things that matter" to "I have the power and expertise to make things that matter."

Linked Open Data and Community

It is important to note that LOD is not universally applicable to all fields, disciplines, courses, or kinds of assignments. The payoff is clear for assignments designed around the concrete and known *who, what, when, where* that LOD is particularly well-suited to connecting. However, as Lynn Ramey cogently noted, the connections between the work of literary scholarship and LOD may be less clear. Discussing ways that LOD might be useful for pedagogy in literary courses, the group hypothesized that literature courses focusing on topics like travel literature and/or crusades literature might benefit from hands-on assignments that map references to specific places (real or imaginary) using LOD identifiers.[53]

In a similar approach, LOD can be meaningfully integrated into literature (and other) courses in fields that do not have as clear connections and payoffs associated with LOD. As libraries move metadata from discrete online catalogs onto the web as linked open data they require a robust LOD ecosystem to thrive. Having students

---

[51] http://orcid.org/.

[52] http://www.vivoweb.org/.

[53] LOD-as-a-site-for-discovery, however, does remain relevant in courses on medieval literature. For example, one might show students how LOD identifiers connect diverse sources and resources, on, say, Prester John's kingdom. It is completely possible to use LOD for fictional places just as easily, for which see, http://dbpedia.org/page/Narnia_(world), which is the DBpedia URI for "Narnia."

meaningfully incorporate LOD into their assignments need not only benefit medievalists or medievalist scholarship. In partnership with archivists and librarians, we can use our assignments to slowly build up the rich data sets on which LOD depends. Even if that specific work does not have an immediate additional value for Medieval Studies or a particular scholar's work, it may be extremely beneficial for our partners in libraries and archives, and thus still have the real-world-impact value that students benefit from and crave.

The questions of LOD in medieval and early modern pedagogy are not just about how LOD can be leveraged to help our particular research interests and sub-fields. If libraries are embracing LOD, *scholars must consider what they can and need to do to support the library in this endeavor such that we all (who individually feel so embattled) thrive*. How can undergraduate students and meaningful student work become a central and publicly-valued part of this network of mutually-enriching intellectual community?

LOD Infrastructure Needs and Digital Pedagogies

1) Tools and tagging methods: Adam Rabinowitz's assignment shows one way that students can begin, now, to engage with and create LOD in classroom assignments. Another simple and useful way to leverage LOD for productive assignments in medieval courses is to create assignments that send students to image archives and have them tag items with LOD identifiers associated with a particular place or time period the class is currently studying. (Think of the tagging capabilities of sites like Flickr, modified.) For example, an instructor might send students to find and tag digital images of physical objects associated with Alexandria or Athens using LOD place identifiers from Pleiades.[54] Another instructor might have students find and tag medieval manuscript images using LOD identifiers from Periods, Organized (PeriodO). Unfortunately, it is not yet possible to harvest students' tags so that students' annotations can be represented in RDF. Some tools, such as Hypothesis, allow tagging, but do not always allow the data to be scraped and repurposed.[55] Assignments being created *now*, however, could still have a tagging component

[54] http://pleiades.stoa.org/home.

[55] For example, the Perseids platform has been able to translate Hypothesis annotations into URIs: see https://sites.tufts.edu/perseids/news-and-updates/journey-of-the-hero/; http://perseids.org/sites/joth/#index and here for code: https://github.com/perseids-project/hypothesis-client. We thank Adam Rabinowitz for the reference.

in anticipation of the time when scraping and reuse is supported (ideally, only 2 or 3 years away). A paleography/codicology course, for example, could have the requirement of including LOD identifiers for place and author, where possible, in students' final digital projects so that, when the technology develops, we have rich data to work with and work from.

2) Community-building: LOD is not the work of scholars *or* librarians *or* technologists. Instead, it is an idea and an opportunity that will only bloom in ongoing shared conversations between all of these groups. If there are assignments that professors can create that use LOD to help librarians' goals, then that too is a mode of experiential learning that
    a) Uses a shared investment in undergraduate pedagogy to forge and maintain good relationships between faculty, subject and metadata librarians, and archivists, and
    b) Gives students real world clients (libraries and archives) and, therefore, a sense of their work and college education as devoted to creating *work that matters*.

In the evolving academic environment it is important for us to understand that these relationships must be ones of mutual support and not of one-way service. If we create semantic links useful to research and teaching, we need to maintain an open dialog among instructors, librarians, and technologists about how we develop our LOD and how to meaningfully engage undergraduate students in this work.

The greatest element resulting from much of our discussion was a shared commitment, across disciplines and pedagogical approaches, to getting students to see themselves as active knowledge producers, rather than passive containers for received and inherited knowledge. An understanding of and role in the creation of LOD—supported by laddering designed to meet the needs of the diverse student populations of specific institutions is one tool that can help students become more information literate, able to critically assess their information sources, and more responsible for creating new knowledge and new tools for their chosen fields of inquiry.

Peer-Review and Publications-Related Data Group:

This group discussed data creation and reuse, the publication cycle, peer-review, and open-access as well as its applications in medieval studies.

Group Lead: Jodi Reeves Flores (Arizona State University)

Gabriela Curie (University of Minnesota)
Eric Kansa (Open Context)
Lilla Kopár (The Catholic University of America)
Ece Turnator (University of Texas at Austin)

DISCUSSION POINTS:
(Note: These bullet points were used as discussion primers; not all points were discussed in detail during the actual discussion)

- What was the last thing you…
    - Published? (and was it open access?)
    - Put in a digital repository (curated)?
- Where are we with respect to OA (Open Access) in Medieval Studies?
- Where do we want to be?
    - What would you publish and make OA?
    - How do you see other people using your data? (And how does this make you feel?)
- What are the publication and assessment needs for datasets, images, articles, books?
- Do you trust an open access publication? Do you trust a repository? What needs to be in place for there to be trust in the OA process?
- Peer review of digital projects and OA works—who will do it, who will bear the costs?
    - Do your institutions have OA payment systems set up? What are they?

The discussions of this group can be summarized under two headings:

1) **Digital publications**: The group underscored the need to distinguish the different processes and skills required to publish a monograph, an article, or curated data sets (with and without peer-reviewed versions of these different kinds of publications).

Further, the group emphasized the importance of the timing and managing the quality of digital publications and focused on having an initial realm of privacy, or a "safe space." Creating publicity and wide access in early phases of data curation and project development was seen as a potential impediment to productivity, and possibly harmful. In light of these concerns, the group elected to consider the peer-review and publication processes for digital projects rather than that on digital publications.

2) Digital Projects: The infrastructural needs and different nature of digital projects have caused a shift in both the publication and evaluation of their research outcomes.[56] Therefore, a system needs to be established to assure that the prestige of the received grants for project funding does not end up replacing actual scholarly value.

In order to prevent value assignments of digital projects by size of received grant, there is a need to establish some objective criteria, such as:[57]

1) Does the digital project follow best practices in user centered design?
2) Is it findable?
3) Is it accessible?
4) Is it available on the devices its users require?
5) Is it innovative/does it solve problems in a new way or with a new technology?
6) Does it get press?
7) Is it doing an important service to its field(s)?
8) Have there been presentations/publications about it?
9) Does it make hard-to-access, fragile, or endangered resources available?
10) Does it have a robust preservation plan?
11) Does it facilitate important research for distributed researchers that would otherwise have to travel?
12) Has it attracted faculty partnerships, or have faculty partnerships been incorporated into academic curricula?
13) Do a lot of people use it (relative to the size of the primary identified audience)?
14) Are other people citing it/blogging about it?

[56] See n. 2 above.

[57] These robust assessment criteria are proposed by Jade Anderson Diaz, Project Manager, Technology Resources, UT Austin. The list is not in any particular order of importance.

15) Does is spark other research?[58]
16) Does it get grant funding?
17) Does it build on previous research or make it more widely available/findable?

Other means of measurement are possible and need to be collectively discussed by scholars. Different fields will have different requirements and focuses. Professional organizations, such as the American Musicological Society or the Association of Art Historians, could rally their scholarly communities around creating objective criteria for evaluating digital projects. This could instigate individual institutions and scholars in specific fields to take action. Established scholarly societies and associations would offer a valuable service by documenting the best practices and establishing a blueprint for evaluating digital projects in their own fields. Because the way individual scholarly communities receive and review digital projects is key, the involvement of these supra-institutional scholarly societies becomes invaluable in establishing the required principles and best practices in evaluating digital projects.

In short, the establishment of a robust and reliable **review process** is necessary. The mechanism for that process involves linking the reviews of the project to the project itself and publishing them altogether in a digital publication platform, such as the *Journal of Digital Humanities* or the *Digital Humanities Quarterly* (DHQ), or other publishing venues. These actions help with dissemination. *Speculum*, *Digital Philology*, *The Medieval Review*, and *the Ethnomusicology Journal* are just a few potential medieval-specific outlets for the evaluation of digital medieval projects. Alternative media, such as Chuck Jones's The Ancient World Online (AWOL) and the Medieval Electronic Scholarly Alliance, can also be used for dissemination.

The incorporation of a new publishing model that builds upon the current publication model, rather than completely dismantling it, could be one way of solving the "publication with peer-review" problem. Alternate solutions can be field-specific. One example of this type of solution is the initiative started by the University of Colorado at Boulder. Its university press is working together with several of its peers for the purpose of exploring "ways to deliver data- and illustration-rich digital editions of cutting-edge archaeological research."[59] Alternatively, it is possible to create new publishing models like the one built by Harold Varmus, et. al., in the

[58] One way of measuring the success of a digital project could be the number and quality of the follow-up projects that get inspired or informed by a specific digital project. These new projects can either build on content/data or on methodologies and best practices established by the original project.
[59] See http://uanews.org/story/ua-press-receives-second-mellon-grant-2009.

Public Library of Science (PLOS One) and Digital Humanities Quarterly by Daniel Cohen et. al.

Action Plan: Lilla Kopár and Dorothy Porter will contact the editors of *Speculum* to find out if a separate section could be created under the existing "Reviews" section in *Speculum* to regularly publish reviews of digital projects/databases/works relevant for medieval studies. It could be possible to have a DH review editor on the *Speculum* board, and have a group of DH advocates offer to establish guidelines for the reviews. A traditional forum and mode of evaluation may help bridge the gap between traditional and digital scholarship, promote DH projects, and serve as measurable outcome/reception of digital scholarship (for tenure, promotion, etc.). This might be a small step but a very important one towards incorporating the peer-review of digital projects into *Speculum* or other medieval journals' established workflows so that, eventually, these projects could be used in departmental hiring processes.

What can individual scholars do to increase the visibility of digital projects? We can simply refer to them in our work! Gabriela Currie mentioned that she uses the 3D reconstructions of Greek amphitheaters in her research on the uses and applications of acoustics. She pointed out that she uses them as she would a published scholarly work: "each one of [the 3D reconstructions] are interpretive and I do cite them."

Challenges to Changes in Peer-review and Publications in Humanities Projects:

- Humanities evaluation systems are often counter-collaborative. Scholars are evaluated based on their individual merit. A scholar's monograph, incorporating her original idea, is what secures a tenure-track job; edited volumes or being manager of a digital collaborative team are valuable, but insufficient in comparison.

- Given this environment the evaluation of digital projects and incorporation of them into the tenure process falls mostly on the senior faculty, especially those who have participated in collaborative digital projects. They can convincingly argue for the value of a project and verify how much time commitment a digital project requires in front of a hiring committee.

- Not all humanities fields are quantifiable to the same degree. Without losing the perceived value of small data specific to each humanities field, there is a need to come up with a way to objectively assess the quality of the work in humanities fields.

- Including scholars who work outside of universities or other research institutions is important to establish a system where their valuable input/feedback gets incorporated into mainstream scholarship.

- The group also emphasized the need for open access online scholarly publication platforms. Scalar was brought up as a well-supported example of an online open source publication platform. Even with a platform like Scalar, however, the means to incorporate, assign due credit to the editorial work, *and* to integrate a robust peer-review process into online publications remain significant parts of the puzzle that await a solution. Editorial boards and content specialists who complete the peer-review are still needed, even though there are journals like *Digital Philology* or *Digital Medievalist* who are taking significant steps in this direction.

- Citation metrics: Current evaluation and citation metrics are not reliable and/or fair. Scholarly work on popular topics (e.g., articles/books on Chaucer/Shakespeare/Vikings) accrue more citations by dint of that popularity. Given an article on those topics that has the same intellectual merit as another scholar's valuable work straddling both mathematics and theory of music within the medieval context, the former will appear to be a more important contribution. Thus, simply counting citation numbers or page views ends up narrowing the field to a *de facto* canon of popular themes. There is a real need for developing fair, smart, non-homogeneous metrics.

- Additionally, more and equally valid career tracks that allow for alternative ways of contributing to scholarship for humanities graduates need to be developed and articulated within the bounds of academic freedom and increased access to quality scholarly work.[60]

[60] William Pannapacker, "Cultivating Partnerships in the Digital Humanities." *The Chronicle of Higher Education* 13 May 2013 http://chronicle.com/article/Cultivating—Partnerships--in/139161/.

WORKSHOP PARTICIPANTS:

Christopher P. Atwood is Associate Professor of Mongolian Studies in the Central Eurasian Studies Department at Indiana University Bloomington. He has published numerous books and articles on Mongolia including *Young Mongols and Vigilantes in Inner Mongolia's Interregnum Decades, 1911-1931* and *Encyclopedia of Mongolia and the Mongol Empire*. His current research focuses on using the tools of philology and source criticism to place Mongolian-language history writing during the Mongol empire in its full Eurasian context.

Alexandra Bolintineanu has served as a CLIR Postdoctoral Fellow in Data Curation and Medieval Studies at University of Toronto, jointly affiliated with the Centre of Medieval Studies and the Jackman Humanities Institute. She researches Old and Middle English spatial poetics of wonders and provides technical and data curation services to digital medieval studies research projects. Email address: alexandrabolintineanu@gmail.com.

Adam Brin is the Director of Technology at The Digital Archaeological Record (tDAR). As Director of Technology, Adam is responsible for the design, development, and maintenance of tDAR. He has spent his career at the intersection of cultural heritage and technology, providing consulting and programming services to museums, libraries, and software companies. His projects include work with NASA, the Internet Archive, and Luna Imaging to create a centralized database of all of NASA's online images; work with the David Rumsey Map Collection; and work with the University of California and OCLC on their Next Generation Melvyl Project. Adam specializes in developing simple and elegant services for complex projects with unique metadata and technical challenges.

Sayeed Choudhury is the Associate Dean for Research Data Management and Hodson Director of the Digital Research and Curation Center at the Sheridan Libraries of Johns Hopkins University. He is a member of the Executive Committee for the Institute of Data Intensive Engineering and Science (IDIES) based at Johns Hopkins. He has been a member of the National Academies Board on Research Data and Information, the ICPSR Council, the DuraSpace Board, Digital Library Federation advisory committee, Library of Congress' National Digital Stewardship Alliance Coordinating Committee, and Federation of Earth Scientists Information Partnership (ESIP) Executive Committee. He has been a Senior Presidential Fellow with the Council on Library and Information Resources, a Lecturer in the Department of Computer Science at Johns Hopkins, and a Research Fellow at the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign. He is the recipient of the 2012 OCLC/LITA Kilgour Award. He has testified for the Research Subcommittee of the Congressional Committee on

Science, Space and Technology. Choudhury has oversight for data curation research and development at the Sheridan Libraries at Johns Hopkins University. His full biography may be accessed here: http://www.educause.edu/members/sayeed-choudhury

Melanie Cofield is the Metadata Coordinator in the Cataloging and Metadata Services department at the University of Texas Libraries (UTL). Her efforts to facilitate UTL metadata practice and policy in support of digital curation and interoperability are informed by engagement with UT staff, faculty, and students across campus to identify metadata needs, prospective projects, and training opportunities. Since Melanie joined UTL in September 2014, she has focused on learning about linked data for libraries, the future of resource description, and the Semantic Web.

Gabriela Currie is Associate Professor of Musicology at the University of Minnesota. Her research interests and publications concern medieval music theory, the intersection between musical and scientific thought in the early- and pre-modern eras, music iconography in pre-modern Eurasia, and travel accounts as early ethnographies of Byzantine, Balkan, and Ottoman musical traditions. Currently she is working on several projects involving pre-modern Eurasian music iconography and preparing a book on the intersection of late scholastic natural philosophy, mathematics, theories of sound, and musical cosmologies in the works of Nicole Oresme.

Jee-Hyun Davis has been Head Librarian of Cataloging & Metadata Services at the University of Texas Libraries since 2010. She has held various positions in academic libraries over the last fifteen years. Davis has a wide range of research interests including metadata policy, linked open data in libraries, workplace transformation, leadership and management, and diversity in academic libraries. She is a fellow of the 2010 Minnesota Institute as well as 2013-2014 ARL Leadership and Career Development Program.

Matthew Evan Davis is currently the Council for Library and Information Resources/Mellon Postdoctoral Fellow in Data Curation for Medieval Studies at North Carolina State University. There, he works as part of the team on two similar projects—the *Piers Plowman Electronic Archive* and the *Siege of Jerusalem Electronic Archive*, as well the Manuscript DNA project and the Medieval Electronic Scholarly Alliance, an aggregator and discussion space for digital scholarly and cultural heritage work regarding the Middle Ages.

Maria Esteva is a Data Curator at the Texas Advanced Computing Center at the University of Texas at Austin where she works with researchers from diverse domains helping them to model their research workflows and map lifecycle data curation activities to cyberinfrastructure resources. She conducts a significant collaboration with the Institute of Classical Archaeology at UT Austin in the development of their collection architecture which includes end-to-end data

organization, sharing, exchange, publication, and archiving functions. This particular work has been presented at IS&T Archiving, Digital Curation, Digital Humanities, Quantitative Methods in Archaeology, and at the Society of American Archaeology annual conferences. Dr. Esteva's latest work focuses on big data curation. Her latest publications and presentations are about mass video curation using image and video quality control algorithms, and design and implementation of multi-tasking data management workflows on High Performance Computing resources.

Monica Green is Professor of History at Arizona State University. Together with Florence Eliza Glaze and other scholars, she is compiling a list of all Latin medical manuscripts from the "long 12th century" (ca. 1075 to ca. 1225), a project she hopes to make into an open-access database for researching the history of medicine in the Mediterranean and Europe. Together with Kathleen Walker-Meikle, she is working on a "born digital" edition of the late 11th-century *Antidotarium magnum*, a massive book of remedies that incorporates the newly introduces *materia medica* from the Islamic world. Finally, together with a team of historians, anthropologists, and biologists, she recently brought out the collection, *Pandemic Disease in the Medieval World: Rethinking the Black Death*, and she is now thinking what Big Data might be and do for a larger, multi-disciplinary project that would create a rich historical epidemiology of the Black Death.

Jennifer Hecker is Digital Archives Access Strategist at University of Texas Libraries. She brings her experience as a manuscript archivist to bear on digital access issues at the University of Texas Libraries. She is excited about the potential LOD has to increase discoverability, help researchers access library and archives collections, and engage students and scholars in resource description and metadata augmentation activities.

Geraldine Heng is Perceval Associate Professor of English and Comparative Literature, Women's Studies and Middle Eastern Studies, at the University of Texas at Austin. The author of *Empire of Magic: Medieval Romance and the Politics of Cultural Fantasy* (Columbia UP, 521 + xii pp., 2003, 2004, 2012), she is currently completing *The Invention of Race in the European Middle Ages* and *England and the Jews: How Religion and Violence Created the First Racial State in the West*. Heng publishes in *PMLA*, *differences*, and *MLN*, among other journals, and recently co-edited with Lynn Ramey a special issue of *Literature Compass* on the Global Middle Ages. She founded and co-directs the Global Middle Ages Project (G-MAP).

Leif Isaksen is Associate Professor in Digital Humanities at the University of Southampton. His research focuses on the application of spatial and Linked Data technologies in the Humanities and the early development of geographic thought. He is Director of the Pelagios 3 project, and an Executive Committee member of the European Association for the Digital Humanities.

Eric Kansa is the Program Director for Open Context, an open data publishing venue for archaeology (http://opencontext.org). He develops workflows and standards to guide editorial practices and peer-review processes to improve the quality, discoverability, and usability of archaeological data sets. His research interests explore web architecture, service design and how these issues relate to the social and professional context of the digital humanities and social sciences. He also researches policy issues relating to intellectual property, including text-mining and cultural property concerns, and actively participates in a number of Open Science, Open Government, Linked Open Data, and scholarly user needs initiatives.

Thomas Kealy is Associate Professor of Humanities at Colby Sawyer College. Among the topics he has taught are Shakespeare, world literature, and composition. is an expert in world literature and writing, focused on science and nature; the relationship between literature and science; and wilderness education. Among Professor Kealy's numerous awards are a Fulbright fellowship, a doctoral fellowship for international leadership, and the Jack Jensen Award for Excellence in Teaching.

Ward Keeler is a cultural anthropologist with research experience in two Indonesian societies (Java and Bali) and in Burma. His work has ranged over a number of topics, most notably performing arts, literature, gender and sexuality, and language and culture. His most recent research focused on Burmese social relations and their links to Buddhist ideals, based on fieldwork living in monasteries in Mandalay.

Lilla Kopár is Associate Professor of English and Director of the Center for Medieval and Byzantine Studies at The Catholic University of America. Her research focuses on the visual culture of Anglo-Saxon England, in particular on stone sculpture, on myth and religion in Viking Age England, and on stone monuments and inscribed objects in early medieval northern Europe. Her publications include *Gods and Settlers: The Iconography of Norse Mythology in Anglo-Scandinavian Sculpture* (Brepols, 2012) and articles on early medieval art and iconography of the British Isles, Scandinavia, and Greenland, Anglo-Saxon runic inscriptions, and on Old English time words. Kopár is co-director and PI of the NEH ODH funded international collaboration Project Andvari. She has been a visiting scholar at Cambridge, Leicester, Durham, Göttingen, Munich, and Eichstätt, and received several grants and fellowships.

Colleen Lyon is the Scholarly Communications Librarian at the University of Texas at Austin. She is also the manager of the UT Digital Repository—an online archive for the scholarly work being created at UT Austin. She is interested in discovering ways the library can support scholars and improve access to scholarly works.

Roger L. Martinez-Davila is Assistant Professor of History at the University of Colorado at Colorado Springs. From fall 2008 to spring 2010, he served as the Burton Postdoctoral Fellow at *St. Joseph's University* (Philadelphia, Pennsylvania) and earned his Ph.D. in May 2008 from the Department of History at the *University of Texas at Austin*. Dr. Martinez specializes in the study of intercultural relations and how group and individual identities hybridize. He is a scholar of medieval and early modern Spain, religious minorities and religious converts in Spain (in particular, Sephardic Jews and conversos), and their Spanish trans-Atlantic migration to Mexico and Bolivia. Dr. Martinez continues an active research agenda, especially in the area of applying digital tools to the study of medieval and early modern inter-religious Jewish, Catholic, and Muslim coexistence. Currently, Dr. Martinez is the project director for an emerging digital humanities project titled, *Revealing Cooperation and Conflict: An Integrated Geovisual and Transcription Project for Plasencia, Spain (circa 1390-1450)*. The Revealing Cooperation and Conflict project will invigorate the humanities and public's imagination by creating a visually-compelling, data-robust, and historically-lush digital world known as *Virtual Plasencia*.

Lee Mordechai is a graduate student at Princeton University. He is mainly interested in the Eastern Roman (Byzantine) Empire, which he tries to approach using wide chronological, geographical and cultural perspectives. His doctoral dissertation deals with the changing nature of rulership and the emperor's figure in the turbulent political history of the Eastern Roman Empire during the eleventh century while rethinking the traditional scholarly narrative of this period. Lee is also the director of the FLAME (Framing the Late Antique and early Medieval Economy) project, which aims to analyze and interpret the transformation of the economy during the late antique and early medieval periods using the surviving coinage as a proxy.

Kent Norsworthy is the Digital Scholarship Coordinator at LLILAS Benson Latin American Studies and Collections at the University of Texas at Austin. He holds a dual appointment in the College of Liberal Arts and the University of Texas Libraries. His current work is focused on integrating digital resources, tools and methods in research, and in the Latin American Studies classroom at the undergraduate and graduate level. He also provides stewardship for a large portfolio of Latin American Studies digital initiatives, including Primeros Libros de las Américas, a collaborative project that seeks to digitize the first books printed in the Americas prior to 1601, and the Guatemalan National Police Historical Archive, which provides online access to over 10 million pages of human rights records.

Timothy Pauketat is Professor of Anthropology and Medieval Studies at the University of Illinois at Urbana-Champaign. He is best known for his investigations at Cahokia, the major center of ancient Mississippian culture in the American Bottom area of

Illinois near St. Louis, Missouri. Pauketat has concentrated research on Cahokia, the center of the large, regional Mississippian culture that extended throughout the Mississippi Valley and tributaries. He has excavated at its grand plaza and the surrounding platform mounds. Pauketat has used research from contemporaneous archaeological sites to formulate a comprehensive, large-scale picture of the Mississippian world. He is interested in investigating such questions as the emergence of the civilization. He has studied beyond his specialty area to find the unique factors that contribute to his "historical processual" analysis. Reexamination of data and artifacts to discover new or previously ignored information is another highlight of Dr. Pauketat's work.

Dorothy Porter is the Curator of Digital Research Services in the Schoenberg Institute for Manuscript Studies, University of Pennsylvania Libraries, and Medieval Electronic Scholarly Alliance (MESA). As curator, Dot works with Penn scholars in exploring new methods of research in the humanities, particularly the application of digital technologies to textual analysis and the electronic dissemination of humanities research. She holds Master's degrees in Medieval Studies and Library Science and started her career working on image-based digital editions of medieval manuscripts. She has worked on a variety of digital humanities projects over a decade-long career, focusing on materials as diverse as ancient texts and Russian religious folklore, providing both technical support and scholarly expertise.

Adam Rabinowitz is an Associate Professor in the Department of Classics and Assistant Director of the Institute of Classical Archaeology at The University of Texas at Austin. He holds his Ph.D. (2004) from the Interdepartmental Program in Classical Art and Archaeology at the University of Michigan. He is a field archaeologist with a focus on ancient social relations as expressed through commensal practices and colonial interactions. His interest in the use of digital platforms for archaeological documentation and publication began during his work at the Roman site of Cosa in the 1990s and intensified in the course of excavations in the South Region of the Greek, Roman, and Byzantine site of Chersonesos in Crimea in the mid-2000s. Since then, in the course of his preparation of the South Region excavations for publication, he has also become involved with long-term archival preservation and the digital dissemination of rich contextual datasets. He is involved in several digital humanities projects related to the linking and visualization of information about the Classical past, including Pleiades (http://pleiades.stoa.org), a spatial gazetteer of ancient places; GeoDia (http://geodia.laits.utexas.edu), an interactive spatial timeline of Mediterranean archaeology; Hestia2 (http://hestia.open.ac.uk/), a narrative time-map of the *Histories* of Herodotus; and PeriodO (http://perio.do), a gazetteer of scholarly definitions of archaeological, historical, and art-historical periods.

Lynn Ramey is Associate Professor of French at Vanderbilt University where she specializes in Medieval French literature and media studies. Ramey is the author of

Christian, Saracen and Genre in Medieval French Literature (Routledge, 2001) and *Black Legacies: Race and the European Middle Ages* (Florida, 2014), and co-editor with Tison Pugh of *Race, Class and Gender in "Medieval" Cinema* (Palgrave, 2007). She is currently working with recreations of medieval literature and culture in video games.

Jodi Reeves Flores holds a Ph.D. in Archaeology from the University of Exeter. She received a CLIR/DLF Postdoctoral Fellowship in Data Curation for the Sciences and Social Sciences, which was hosted by Arizona State from 2013-2015. While at Arizona State, she served a dual appointment with ASU Libraries and the Center for Digital Antiquity. During her fellowship, she developed seminars and workshops on data management and curation, supervised the curation of digital archaeological data into tDAR (the Digital Archaeological Record), and collaborated on digital humanities projects. Jodi is also on the Secretariat for EXARC, the ICOM Affiliated Organization representing archaeological open-air museums, experimental archaeology, ancient technology, and interpretation.

Tamsyn Rose-Steel is a CLIR/Mellon Postdoctoral Fellow in Data Curation for Medieval Studies, with a joint appointment in the Digital Research and Curation Center of the Sheridan Libraries and the Department of German and Romance Languages and Literatures, at Johns Hopkins University. Tamsyn gained her Ph.D. from the University of Exeter studying citation and allusion in 14th-century French motets. She currently works with JHU's Digital Library of Medieval Manuscripts to develop their online capabilities, and carries out research and teaching in fourteenth-century French music and literature. She is Principal Investigator on the project APRICOT, which is producing a pedagogical hub for teaching medieval topics. Additionally, she is Associate Editor for the complete works' edition of Guillaume de Machaut and has published articles on the medieval motet, citation, and games in medieval literary culture. Email address: tsteel2@jhu.edu.

Ece Turnator received her Ph.D. in Medieval (Byzantine) History from Harvard University in 2013. Her dissertation is an interpretation of 13th-century Byzantine economy through an analysis of archaeological (coins and ceramics) and textual evidence. Since September 2013, she works as a CLIR/Mellon postdoctoral fellow at The University of Texas at Austin in medieval data curation, studying and learning about digital humanities, best practices for data curation and visualization, in addition to teaching and researching in her area of expertise. Her main research interests include world economic history and material culture. Email address: e.turnator@austin.utexas.edu.

Karen Wickett is Assistant Professor at the School of Information, University of Texas at Austin. Her research is on the conceptual and logical foundations of information organization systems and artifacts. She is most interested in the analysis of common concepts in information systems, such as documents, datasets, digital objects, metadata records, and collections. The need for logically consistent

accounts is a pressing issue in digital environments, especially as semantic technologies (computer-processable knowledge representation languages such as RDF, OWL, and SWRL) become more commonplace for digital library and curation systems.

Michael Widner joined Stanford University Libraries in 2012 as the Academic Technology Specialist for the Division of Literatures, Cultures, and Languages. He collaborates with faculty on digital humanities and instructional technology projects, organizes lecture series for the Digital Humanities Focal Group, and teaches courses on digital humanities. Some of the projects on which he works include: Lacuna Stories, a collaborative reading and annotation environment for humanities courses; Decoding Marine LePen's Rhetoric, a text-mining and visualization project; and Performing Trobar, a site for multimodal learning centered on medieval troubadour lyrics. He is the Project Director for Bibliopedia, an NEH-funded platform for the organization, visualization, and sharing of digital scholarly materials that joins researcher needs with library concerns for consistent metadata, preservation, linked open data, and reusability. Widner received his Ph.D. in English from the University of Texas at Austin in 2014, where his focus was medieval English and French literature, genre theory, and cognitive psychology. Email address: mikewidner@stanford.edu.

Bridget Whearty is the Council for Library and Information Resources Postdoctoral Fellow in Data Curation for Medieval Studies at Stanford University, 2013-2015. In this role, she has co-taught paleography and codicology courses, founded and run a medieval manuscripts club for undergraduates, and worked as the data curator for *DMS-Index*, an interoperable index of digitized medieval manuscripts held at institutions around the world. Starting September 2015, she will be Assistant Professor of Medieval Studies and Digital Humanities at Binghamton University, in the SUNY system. Email address: bwhearty@stanford.edu.

APPENDIX:

1. Workshop Program
2. Presentation Abstracts
3. LOD Overview

LINKING THE MIDDLE AGES WORKSHOP
University of Texas at Austin: May 11-12, 2015

The **workshop team created a SURVEY to be completed by participants in advance of the workshop. Questions aim to identify what participants think medieval data is, how it is acquired, used, re-used, and disseminated:**

https://utexas.qualtrics.com/SE/?SID=SV_9KzI0Epf3BYSNRb

LINKING THE MIDDLE AGES WORKSHOP PROGRAM

MONDAY MAY 11, 2015

7:30-8:45 Breakfast **@ Austin Folk House, where out of town participants are staying. Breakfast is for them only.**

9:00-9:30 Introduction **Workshop—Outline funders, goals, participant backgrounds, and objectives by Geraldine Heng and Ece Turnator with a mini overview of linked open data and how it can be applied to Medieval Studies.**

9:30: Presentations

9:30-9:50 Maria Esteva, "The importance of 'metadata ready data' and the challenges of implementing metadata in long and evolving research projects." ABSTRACT
9:50-10:15 Questions & Comments.
Discussion facilitators: Michael Widner (Stanford U) and Dot Porter (UPenn)

10:15:-10:35 Leif Isaksen (via skype), "Linking up with Pelagios. What do you do with a million links?"
10:35-10:55 Questions & Comments.
Discussion facilitators: Adam Brin (tDAR), Adam Rabinowitz (UT Austin)

10:55-11:10 Global Middle Ages Project: Geraldine Heng, Ece Turnator. ABSTRACT
11:10-11:25 Questions & Comments.
Discussion facilitators: Christopher Atwood (Indiana U), Lee Mordechai (Princeton U)

11:25-11:35 Coffee Break

11:35-12:30 Round table discussion of data fields. Linking between what digitized medieval objects (coins, manuscripts, icons, etc.) and data fields (e.g. date, period, geographic location, etc.) would be immediately useful for research and discovery in Medieval studies?

12:30-2:30 Lunch @ Clay Pit

2:45-3:15 Eric Kansa, "Scholarly Communications and Linked Open Data in Archaeology" ABSTRACT
3:15-3:30 Questions & Comments.
Discussion facilitators: Melanie Cofield (UT Austin), Adam Brin (tDAR)

3:30-3:45 Adam Rabinowitz, "PeriodO. ABSTRACT A gazetteer of period assertions for linking and visualizing data. Why is linking periods important and how do we do that?" See his article "It's about time: historical periodization and Linked Ancient World Data".
3:45-4:00 Questions & Comments.
Discussion facilitators: Alexandra Bolintineanu (U of Toronto), Gabriela Currie (U of Minnesota)

4:00-4:15 Lilla Kopár, ABSTRACT "Linking the Visual World of Early Medieval Northern Europe: Project Andvari"
4:15-4:30 Questions & Comments.
Discussion facilitators: Roger Martinez (U of Colorado, C. Springs), Lynn Ramey (Vanderbilt U)

4:30-4:45 Coffee Break

4:45-5:00 Wrap Up, Further Comments.
Lead: Monica Green (Arizona State U)

Wrap up in bullet points what we have learned from presentations, common challenges, and suggested solutions. We will also map out the correlation between the four assessment areas listed under workshop goals.

TUESDAY MAY 12, 2015

Group 1= Research-related data creation and reuse.
Group Lead: Michael Widner (Stanford U)

Group 2= Teaching-related data creation and reuse; what works best, how do we evaluate our teaching, do we share our teaching methods.
Group Lead: Tamsyn Rose-Steel (Johns Hopkins U)

Group 3= Peer-Review- and Publications- related data creation and reuse; publication cycle, peer-review, open-access and its applications in medieval studies.
Group Lead: Jodi Reeves Flores (Arizona State U)

Each group will leave the last 30 mins of their 2 hour allocated discussion time between 9-11 AM to wrap up their discussion and prepare a bullet point report of their discussions to be shared with the other groups in the afternoon.

12:00-1:30 Brown bag lunch

Optional 11:50-1:30 at The Harry Ransom Center: Manuscript Tour. Meet at 2nd floor reception desk 11:50. [Carta ejecutoria. Spain. Dispute Settlement. Valladolid (1534); Treatises with medical recipes; 1256 Charter of Henry III; Antiphonal Leaf (14-15th Century); 15th-Century Venetian Portolan; Cardigan Chaucer (ca. 1450)]

Optional noon-1:30 Linked Open Data Lunch at The Perry-Castañeda Library 3.120
Reconvene at 1:30 in PCL 1.124
1:30-2:30 20 mins. lightning presentations by each group (1:30-1:50; 1:50-2:10; 2:10-2:30)

Each group will do 20 mins. presentation of their discussion main points and prepare a bullet point report of their discussions to be shared with the other groups.

2:30-2:45 Coffee Break

2:45-3:45 Wrap up session
Wrap up session (going over the four points listed under "Assessment Strategy").

- Recap of goals, technological limits of what can be done.
- What have we learned that we didn't know before? Discussion of next steps.

Post- Workshop Task Assignment:

Each group will send their notes to Ece after workshop. Due date: May 30, 2015

CLIR Fellows will edit the results by June 5th, final edits by group leads, and assigned participants will be ready by June 15, 2015.

White Paper to CLIR/Mellon June 2015.

Special Thanks:

Linda Abbey (University of Texas at Austin), Alexandra Bolintineanu (University of Toronto), Adam Brin (The Digital Archaeological Record), Gerald Cloud (Harry Ransom Center, University of Texas at Austin), Melanie Cofield (University of Texas at Austin), Sayeed Choudhury (Johns Hopkins University), Matthew Davis (University of North Carolina), Jade Anderson-Diaz (University of Texas at Austin), Monica Green (Arizona State University), Jennifer Hecker (University of Texas at Austin), Leif Isaksen (Pelagios Project), Eric Kansa (Open Context), Krystal Wyatt-Baxter (University of Texas at Austin), Lilla Kopár (Catholic University of America), Colleen Lyon (University of Texas at Austin), Agnes McAlester (University of Texas at Austin), Lee Mordechai (Princeton University), Kent Norsworthy (Llilas Benson Latin American Studies and Collections, University of Texas at Austin), Adam Rabinowitz (University of Texas at Austin), Lynn Ramey (Vanderbilt University) Jodi Reeves Flores (Arizona State University), Tamsyn Rose-Steel (Johns Hopkins University), Gennell Shuler (University of Texas at Austin), Sarah Whitcher-Kansa (Open Context), Bridget Whearty (Stanford University), Karen Wickett (Information School, University of Texas at Austin) and Michael Widner (Stanford University) for raising the quality of the workshop with their support, input, and suggestions.

PRESENTATION ABSTRACTS (IN ALPHABETICAL ORDER):

Maria Esteva, Research Associate/Data Archivist (Data Intensive Computing Unit) Texas Advanced Computing Center, University of Texas at Austin:
"The importance of "metadata-ready-data" and the challenges of implementing metadata in long and evolving research projects"

Metadata is fundamental to the descriptive richness of digital library projects, and the way in which it is implemented in a project is key to applications such as linking data and interoperability with other digital projects. In the context of scholarly digital projects in which research, data collection development, and data publication functions are intertwined, metadata activities are shifting from being a librarian's task to becoming also the responsibility of the research team. In this context, clear understanding of metadata usage and setting goals for its creation can help researchers accomplish metadata in ways that streamline their research process and enhance the use and reuse of their data.

This presentation will focus on metadata planning and the design of workflows for creating rich, interoperable, and extensible metadata frameworks for evolving research projects. Issues of standards, basic and sophisticated metadata creation, metadata automation and its implementation within unique digital collections architectures will be presented so that researchers can prioritize and organize metadata activities that are creatively integrated with their core research pursuits.

Geraldine Heng, Perceval Fellow and Associate Professor of English and Comparative Literature, University of Texas at Austin and
Ece Turnator Council on Library and Information Services, A. Mellon Postdoctoral Fellow in Medieval Data Curation at University of Texas at Austin, Libraries and Department of English, "The Global Middle Ages Project"

Global Middle Ages Project (GMA) is a portal (currently under development) for the study of the global Middle Ages between 500-1500 A.D. intended to serve the higher-education community, students, and scholars of the global Middle Ages as a gateway to new projects on the medieval world as a whole.

Thinking about the medieval world on a global level gave rise to a host of challenges including establishing consistency in structure, content, as well as database-design. The presentation will include, first, a presentation of our current projects and solutions to data interoperability identified. Next, we will focus on possible common data fields to be used on the portal by medievalists who work on or are interested in cross-cultural data.

The presentation will open up to discussion the best practices in serializing the selected medieval data fields in RDF or JSON formats with the purpose of building an API to let researchers harvest them in meaningful ways. This presentation aims to discuss a linked open data approach to issues of interoperability and access in the field of global medieval studies that might be applicable to other humanities fields.

Leif Isaksen, Pelagios Project Director and University of Southampton: "What do you do with a Million Links?"

Isaksen's presentation is available online: http://pelagios-project.blogspot.com/2015/01/what-do-you-do-with-million-links.html

Eric Kansa, Open Context: "Scholarly Communications and Linked Open Data in Archaeology"

Open Context, an online data publishing platform for archaeology, has now taken part in several studies exploring data publishing and collaborative reuse. These studies help identify opportunities and challenges in implementing linked open data methods in archaeological research.

In this presentation, I will provide an overview of Open Context's approach toward "data sharing as publication" and discuss the workflows and resulting research outcomes of using Linked Data in archaeology. One example will focus on large-scale data integration of animal bone (zooarchaeology) datasets that document early domesticated animals in Anatolia. Another example focuses on a comparative study of energetic and labor investments in monumental architecture, based on the interpretation of excavation field-notes and media.

These case studies illustrate how Linked Data can serve in diverse research applications ranging from the statistical comparison of different datasets to the referencing and citation of narrative and image media.They also highlight how meaningful data sharing requires sustained intellectual investment. Preparing the datasets for publication and analysis required significant effort and expertise (archaeological domain knowledge and familiarity with key ontologies), and ongoing iterative cycles of revision and version control to identify and correct errors and improve semantic modeling.

Lilla Kopár, The Catholic University of America: "Linking the Visual World of Early Medieval Northern Europe: Project Andvari."

Project Andvari is an NEH-funded international collaborative project designed to create a free digital portal that will provide integrated access to collections of northern European art and artifacts of the early medieval period (4th–12th centuries). The pilot (in progress) will aggregate data primarily from The British Museum,

Norwich Castle Museum, and Kringla (Sweden). The current presentation will focus specifically on issues of metadata and related technical aspects of the project as a case study.

The nature and structure of metadata for Andvari has been determined by researchers' needs as much as by practical considerations. Decisions were based on an interdisciplinary conversation in which scholars of various fields stated what they envisioned such a resource would include; LIS practitioners offered recommendations on how to best meet user needs based on actual information seeking behavior; and information technology specialists provided insight into what would be possible, given technological constraints.

While the data fields will reflect the needs of researchers who want metadata to provide meaningful insights to pertinent research questions, our metadata will be driven by available metadata from contributing organizations in order to enable automated harvesting and metadata generation. These institutions will in turn be provided enhanced metadata to further improve their collections. One such area will be iconographic content for which the Andvari team is currently developing a unique vocabulary. For authority control of periodization, object classification, and style, the project team will collaborate with owners of existing (or developing) thesauri (British Museum, PeriodO, etc.). Although a general metadata schema has been developed, we continue to map user expectations to technical realities of creating the least amount of manual work possible via automated ingest features (to be designed by IATH).

In technical terms, the project team has determined that a dynamic web interface is necessary for Andvari to be more than just a new face on old data. This will include searches enhanced by metadata and authorities; pre-coordinated searching based on authorities and user specified facets; and unique discovery layers (i.e. map interfaces, "show me more like this," etc.). Platform interfaces and features should facilitate easy and customizable searches while providing a creative research tool that ultimately points researchers back to the original holding institutions. The interface of the Andvari portal will be based on a custom programming of the open access CMS platform Drupal and will include as many features as can be accomplished in the 1.5 year stage II time frame.

Adam Rabinowitz, University of Texas at Austin "PeriodO: a Gazetteer of Period Assertions for Linking and Visualizing Data. Why is it Important to Include Periods in a Linked Data Infrastructure, and How do we do it?"

The last few years have seen major advances in our ability to connect very different datasets related to the ancient world by a few simple common elements. The most important of these elements has been geographic: the addition of a URI from a spatial gazetteer to the name of a site in a database or a text makes it possible to display the

site on a map by extracting coordinates from a gazetteer. Since those who study the ancient world are often looking for material from a particular place, Pleiades and Pelagios have opened a new world of data discovery. But the same scholars also often look for material from a particular time, or a combination of time and place, and it has proven more difficult to deal with time. This is in part because we often use words—the names of periods—rather than absolute dates to describe the chronological aspects of our material. Moreover, although periods usually have at least vague chronological coordinates, these are highly dependent on both place and scholarly tradition, and definitions are often hotly contested. Unlike ancient places, no central gazetteer of periods has emerged to allow us to link periodized or dated records across different datasets.

PeriodO is an attempt to create such a gazetteer. Instead of trying to standardize both the vocabulary and the conceptual identity of archaeological, historical, and art historical periods, it seeks to record and provide unique identifiers for the chronological and spatial *definitions* of periods offered by authoritative sources.

This presentation will briefly describe the project's history and data model, and will then explore some of the possibilities for the discovery of dated or periodized information the PeriodO gazetteer opens up. Ideally, the following discussion will also identify periodizations and chronological authorities relevant to medieval studies that might be incorporated into the gazetteer.

Linked Open Data Overview[1]

Is an Open and Decentralized Medieval Studies Possible?
Background
Glossary of Key Terms and Concepts
The Goal
How it Works (Scholar's perspective)
Infrastructure it Needs (Institutional perspective)
Overview

Is an Open and Decentralized Medieval Studies Possible?[2]

Digital medieval projects (and digital projects in general) suffer from a lack of interoperability because scholars often do not use common standards, ontologies, vocabularies. Even in cases where such standards are carefully picked to serve individual projects, they do not play well or speak to projects written using different standards. Digital-medievalists list in a post-Babel digital world, and it is highly unlikely that everyone, everywhere, will abandon their favorite standards and ontologies to adhere to some universal schema, the same way no one is going to give up the language they speak to recreate everything they have ever produced in a magical universal language.

LOD lets computers work around barriers to communication between digital resources (an online manuscript, online reference to a coin are examples of such "resources") by letting people declare which standard (or ontology, or vocabulary) they are using when they use it, and encouraging others to link to other sites that refer to the same "resource."

Background:

The web was designed for humans to read, not computers. Tim Berners-Lee, the inventor of the World Wide Web, writes:

---

[1] Prepared by Alexandra Bolintineanu, Matthew Davis, Tamsyn Rose-Steel, Ece Turnator, Bridget Whearty, and Michael Widner. We would like to thank Adam Brin, Leif Isaksen, and Eric Kansa for their feedback on this overview.

[2] L. Isaksen, *Archaeology and the Semantic Web*. Ph.D. Dissertation, University of Southampton, School of Electronics and Computer Science, 2011 (available at http://eprints.soton.ac.uk/206421/).

> The Web was designed as an information space, with the goal that it should be useful not only for human-human communication, but also that machines would be able to participate and help. One of the major obstacles to this has been the fact that most information on the Web is designed for human consumption, and even if it was derived from a database with well-defined meanings (in at least some terms) for its columns, that the structure of the data is not evident to a robot browsing the web.

This fact makes it difficult, if not impossible, for discovery and analysis of much of the web, including traditional humanities scholarship. Transforming text-centric information—or even databases of objects—into linked open data makes it possible for computers to use the descriptions of and interconnections among the multifaceted objects of scholarship. Humanities researchers can then use findings garnered by algorithms to drive new research. In short, LOD transforms isolated pieces of humanities scholarship into networks of knowledge, ultimately enabling new forms of research, contextualization, and sharing to emerge.

Glossary of Key Terms and Concepts[3]

*Semantic Web: The **Semantic Web** is an extension of the **Web** through standards by the World Wide **Web** Consortium (W3C). The standards promote common data formats and exchange protocols on the **Web**, most fundamentally the Resource Description Framework (RDF).*

URI: Unique Resource Identifier. Deemed fundamental to the success of semantic web. It is the web address where information about the resource resides.

*RDF Resource Description Framework: A framework, a model, or a diagram that describes a resource on the web through **triples**, i.e. in the form of subject-predicate-object expressions.*

Triples: The subject denotes the resource, and the predicate denotes traits or aspects of the resource and expresses a relationship between the subject and the object. For example, one way to represent the notion "The sky is blue" in RDF is as the triple: a subject denoting "the sky," a predicate denoting "is," and an object denoting "blue."

[3] These entries are adapted from respective articles on wikipedia.org.

Triplestore: RDF triples may be stored in a type of database called a triplestore.

Serialization: Putting descriptions of web resources into a machine-readable format.

The goal

To facilitate the finding and linking of semantically-related data that cannot be related in a traditional database.

How it works (Scholar's perspective)

Suppose you are a numismatist interested in finding data about mints in Ancient Greece. For this research project, you are only interested in mint location and want to use it. Say you start with an online database of coins. In it, each coin entry has that coin's attributes specified (size, metal content, imagery, mint location, find location, find type, etc.). All of this information will be available in a serialized format (=computer readable) and live on a server, perhaps of the institution that holds the coin. This coin, this resource that lives on the web, has a unique identifier: the web address that does not change because the institution that holds the coin provides the resources to make sure that the web address of that coin does not change. Why is this important?

In her case, every time she wants to use a coin, she copies and uses the URI (again, the stable address of the digital resource that lives on the web) and pastes it in her database when she wants to link her work and use information about that specific coin. This process of using URIs and linking with them indicates to computers that "this is the coin I am using and linking to, along with all the information in the URI that is associated with it."

One way of providing researchers the resources to link to is to produce URI-defined gazetteers, which serve as directories for the URIs that identify each resource within that directory.[4] These gazetteers function very much like look-up tables or phone

[4] On historical gazetteers see AAG Seattle and Harvard Center for Geographic Analysis 2011 Symposium on Space-Time Integration: http://gis.harvard.edu/events/seminar-series/aag-2011-special-track-historical-gazetteers.

directories. Continuing with our analogy of mint location, you add a field to the mint location table that is called the "Pleiades_uri" (Pleiades' web address for an online entity, in its case a geographic location), and in it you put the Pleiades identifier for each of the places mentioned. If the URI refers to Athens in Pleiades, it refers to the capital of modern Greece and not to any other city in the world that is named Athens (like Athens, OH, U.S.A). This way the resources on the web have their own unique IDs and are thus disambiguated even if they have the same name.

One could repeat this process for other information types, such as periods, people, climate information, etc., when they become available in a serialized and linkable format.

The point is to literally say to the computer in which you want to be able to parse your data, "This is the [digital representation of] Athens I am referring to here, this is the coin, this is the person, etc."

One of the essential features of LOD is that global agreement on a single standard, ontology, or vocabulary is not necessary. Users are required, however, to state which standard (or ontology, or vocabulary) they are using and linking to when they copy and paste the URI into their own data.[5]

Inserting URIs is simple but it does not allow a **query of** the data that is linked up; for that, programming expertise is needed.

What LOD is NOT good at doing:

Linked Data is not good at asserting negations. For example, LOD does not say "this reference (which is a reference to London, UK) is *not* a reference to London, Ontario, Canada".

Infrastructure it needs (institutional perspective):

First, a decision regarding the kinds of data to put in LOD and the methods of serializing that data (which involves describing what ontologies and serialization schemata are used in putting the data in serialized form) must be agreed upon. For medievalists, place names are already disambiguated via Pleiades.[6] Period definitions

[5] On vocabularies see, Linked Open Vocabularies, hosted by the Open Knowledge Foundation: http://lov.okfn.org/dataset/lov.

[6] For a gazetteer of place names in Arabic from the Classical Islamic World with links to sources, the Encyclopedia of Islam, Pleiades and Wikipedia, see Maxim Romanov's al-Thurayyā:

(i.e. what dates Late Roman Empire corresponds to, according to whom, in what part(s)/countries of the world was there such a period, etc.) will be similarly disambiguated when Periods, Organized (PeriodO) is ready. A low-lying fruit seems to be coins, and ISAW is already serializing that data for ancient coins.[7] Manuscripts are still on the wish list, although work has been underway since the 2011 Stanford Workshop.[8]

Institutional servers, stable URI providers, are the next basic step, and programmers are needed to put data into serialized form (XML, RDF, JSON, KML, Turtle, etc.).

Additionally, Institutions need to get programmers and domain experts together to add and expose the data in a serialized form. In serialized data, one long document repeats each field and value for each coin. The Pleiades URI for the mint location is repeated for each coin that comes from that mint. This serialized form has to be exposed to web-calls on it somewhere. Institutions need to do this. Here is an example of serialized data from Pleiades, defining Memphis (location in Egypt) and linking it with other web resources that use Memphis in the sense defined in Pleiades: http://pleiades.stoa.org/places/736963/turtle.

Other institutions can do what Pleiades is doing for ancient places for manuscripts, coins, medieval people, institutions, etc. The question is, is it worth their time and resources?

---

http://maximromanov.github.io/althurayya/. For an integrated example of historical gazetteers with maps, statistics, census reports and travel writing all accessible through one portal see, http://www.visionofbritain.org.uk/. Description of the project: Humphrey Southall, "Rebuilding the Great Britain Historical GIS, Part 3: Integrating Qualitative Content for a Sense of Place" Historical Methods: A Journal of Quantitative and Interdisciplinary History, 47 (1), 31-44. DOI: 10.1080/01615440.2013.847774.We thank Leif Isaksen and Maxim Romanov for the references and the links.

[7] http://nomisma.org/.

[8] Report of the Stanford Linked Data Workshop, 27 June-1 July 2011, 46: "In the specific domain of digitized ancient, medieval, and early modern manuscripts and in specific support of the work underway to develop the tools and agreements to support interoperability for scholarly functions across silos of digitized manuscripts, Stanford will collect descriptions of manuscripts in URIs. Then, Stanford or another agency will connect individual applications that are showcasing different sets of medieval manuscripts. These projects, the development of interoperability across silos AND the descriptions of manuscripts expressed in URIs, are extensible to many other domains and their digital repositories."

Some researchers can access and use the data provided in serialized form, which can be queried by a query language called SPARQL.[9]

Overview

At the simplest level, linking involves putting a URI into a cell on a spreadsheet, in the row of an item associated with that resource. The difficult but crucial aspect of LOD is creating an ecosystem of scholars, technologists, and institutions that have a long term commitment for programmers to work with domain experts to serialize data and to expose it on the web, and perhaps building an Application Programming Interface to let people harvest it in meaningful ways from whichever institution's server it is exposed. Once the data is serialized and exposed in order for it to be queried, programming knowledge is required. However, even without the ability to query the resources, LOD allows the internet to become an open and semantically linked web of resources, and that, *per se,* is an invaluable feat—perhaps a step for something greater.

In short, LOD is a Web-based technology and has very limited application in non-Web contexts. It is possible to create and use RDF and URIs entirely on a local system but under those circumstances it is not clear if the benefits outweigh the costs. The community that will reap the most from LOD will be that which is willing to put content online.[10]

LOD is only as useful if content is actually accessible. A link to a resource behind a paywall is almost (but not quite) as useless as no link at all. The community that will reap the most from LOD is the one that makes their content **openly** accessible.

LOD is by nature decentralized, which to some extent requires trust in other people's vocabularies, datasets, etc. It is possible to create a private data cloud in which one controls all the content, but then why not just build a database? Those gaining greatest advantage from LOD will be those willing to place trust in building upon resources produced by trusted collaborators.[11]

---

[9] http://www.w3.org/TR/rdf-sparql-query/.

[10] See http://5stardata.info/.

[11] We thank Leif Isaksen for the last three paragraphs above.

Workshop Survey Questions

Primer and Question 1)

This survey aims to assess medievalists' workflows, from note-taking to publication. We aim to understand how medievalists conduct their research and what digital tools, formats, platforms they currently use in their research, teaching, and publications.

NOTE: Please use Firefox or Chrome.

List at least three websites/digital resources you use most often. State what you use them for.

1

2

3

More

Question 2)

What kinds of information do you search for to use in your research/teaching? (Select all that apply).

- Author name
- Work title or keyword in title
- Geographic place / location
- Historical entity (Second Crusade, Yoruba people, Carolingian Empire)
- Historical period
- Historical person
- Material type (ceramics, coins, icons, glass, bronze, textile, musical instruments, etc.)
- Languages(s) written and/or spoken
- Climate
- Fauna/Flora
- Primary industries
- Primary crops
- Dwelling types

Population estimate(s)

Other (if you make more than one entry, please separate entries with a semi-colon)

Question 3)

You use digital resources to... (Select all that apply).

	Frequently	Sometimes	Rarely	Never
Access primary texts	☐	☐	☐	☐
Access online dictionaries	☐	☐	☐	☐
Access monographs and journals	☐	☐	☐	☐
Access digital projects on my area of expertise	☐	☐	☐	☐
Access databases	☐	☐	☐	☐
Find images	☐	☐	☐	☐
Access excavation reports	☐	☐	☐	☐
Access maps and GIS data	☐	☐	☐	☐
Access blogs	☐	☐	☐	☐
Other (please type)	☐	☐	☐	☐

Question 4)

Which one of the below best describes your note taking process(es)? Select all that apply.

- Take notes in a word processing document in my computer
- Take notes in a database, table, or spreadsheet
- Take notes by hand on paper/book/article
- Take notes in a citation management tool
- Other (please type)

Question 5)

Approximately what percentage of your research and teaching materials are digital and what percentage are analog? (Slide for digital material percentage, the rest considered analog).

0 10 20 30 40 50 60 70 80 90 100

Digital research materials

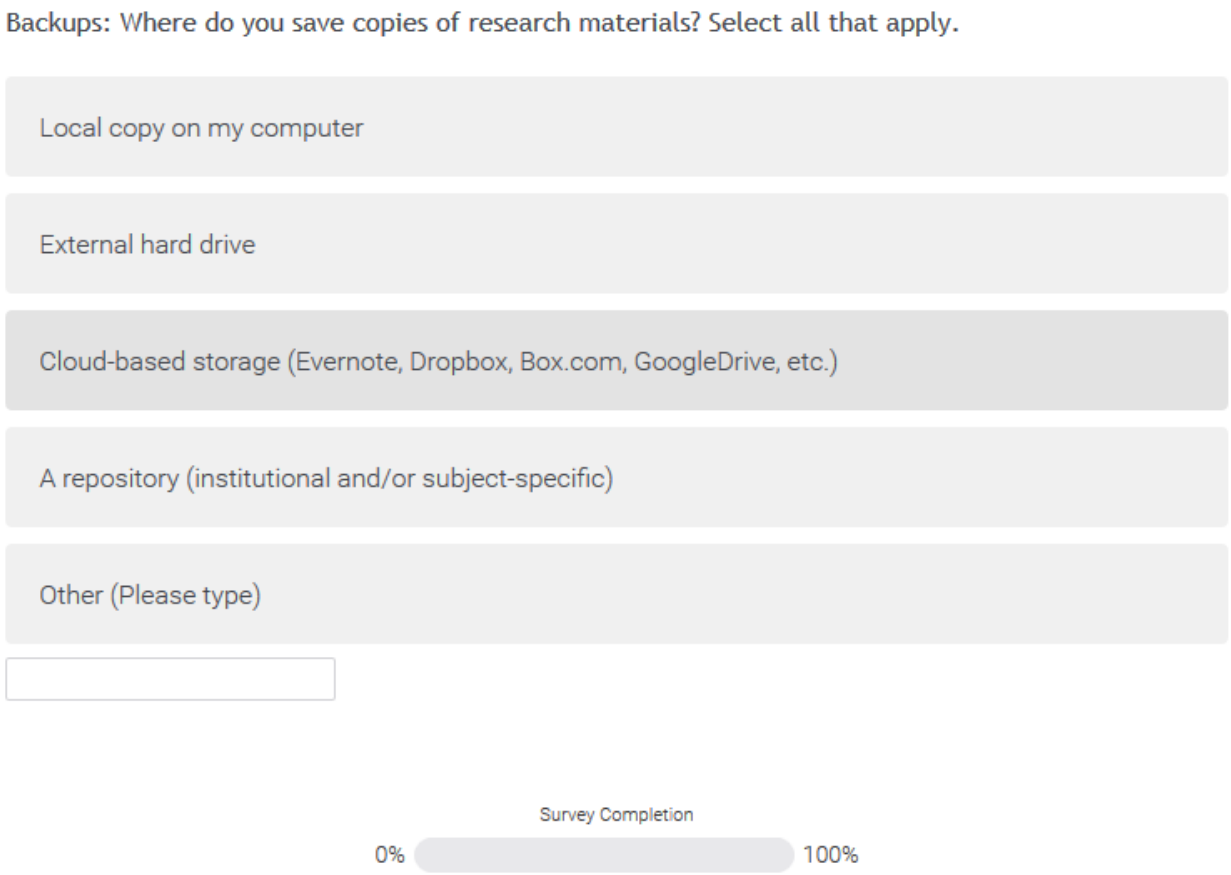
Digital teaching materials

Question 6)

When it comes to note-taking... Select all that apply.

- I keep more than one type of notes (for example, on Zotero but also as notes inside a physical book or a PDF, or email)
- I categorize my notes: by theme, author, title, etc.
- I sometimes cannot access my old notes, because I can't remember where they are.
- I sometimes cannot use my old notes, because they are in a different format from the one I am using now.
- Other (please type)

Question 7)


Backups: Where do you save copies of research materials? Select all that apply.
Local copy on my computer
External hard drive
Cloud-based storage (Evernote, Dropbox, Box.com, GoogleDrive, etc.)
A repository (institutional and/or subject-specific)
Other (Please type)
Survey Completion
0%
100%

Question 8)

Which citation management tools do you use? Select all that apply.

- Zotero
- Mendeley
- EndNote
- Refworks
- Other (please type)
- I don't use any

Question 9)

How often do you use the below software in your research and teaching?

	How often?			
	Frequently	Sometimes	Rarely	Never
Word processing software (Word, OpenOffice/LibreOffice, Pages, Scrivener, Google Docs)	○	○	○	○
Spreadsheets (Excel, Calc, Google Sheets)	○	○	○	○
Text markup, transcription or annotation tools (TEI, XSLT, Oxygen)	○	○	○	○
Mapping Software (e.g. Google Fusion Tables, WorldMap, GIS)	○	○	○	○
Other (please type)	○	○	○	○

Question 10)

Which Operating System(s) do you use often? (Select all that apply.)

- Apple OSX
- Apple IOS
- Android
- Microsoft Windows
- Non-Apple *nix (Linux, BSD, Debian, etc.)

Question 11)

How often do you work on digital projects (online exhibitions, scholarly digital publications, shared databases, 3D projects, etc.)?

- Never
- Rarely
- Sometimes
- Frequently

Question 12)

If you collaborate on digital projects who do you collaborate with? Select all that apply.

- With the IT unit at my institution
- With another institution via a grant
- With my library
- Other

Question 13)

Which DH skills would you like to acquire? Please rank in order of importance.

Text mining

Topic modelling

TEI markup

Network analysis

Data visualization tools

Statistical analyses

Geographic information systems

Database management and design

Content management system (Drupal, Joomla, Wordpress)

Digital exhibit building (Omeka, Neatline)

3D designs and exhibits

Other (please type)

Question 14)

If editing and peer-review was done at no additional cost which research and teaching output(s) would you consider publishing open-access? Select all that apply.

- Books
- Articles
- Research data
- My teaching materials (syllabi, notes, presentations, etc.)
- I would not publish any research/teaching materials in open access form
- Other

Questions 15-17)

Is there anything else you want us to know about your research, teaching, publishing practices?

Your name, institutional affiliation (if any), and email address

You are a... (We are asking this question because we are interested in finding out if there are differences among groups)

Graduate student | Postdoc | Librarian | Researcher/Non-faculty Lecturer | Faculty | Other

Survey Completion

0% 100%