White Paper Report

Application Number: HD-51538-12

The Scholar's Dashboard:
Creating a multidisciplinary tool via design and build workshops

Project Director: Gwen Evans

Award Recipient: Wright State University

Date Submitted: June 12, 2014

1

**Abstract**

Despite challenges including a major shift in project personnel and the dissolution of the depository for which the project was originally intended, the Scholar's Dashboard project was a success in bringing together humanities scholars, librarians and archivists, and technologists to generate functional and technical requirements for digital humanities repositories. Many of these included principles already familiar to digital humanities practice (the ability to visualize or download search results, API's, etc.). On the other hand, the participants showed a greater willingness to consider commercial products and services than is common in the digital humanities community, both for their advanced functionalities and because software-as-a-service and commercial hosting may be more sustainable for many institutions than open-source and/or self-hosted solutions to digital humanities challenges.

**Project Activities**

The NEH funded Scholar's Dashboard to support a series of three themed workshops to lay the groundwork for a new user interface for the OhioLINK Digital Resource Commons (DRC). At the time the grant was awarded (Spring, 2012), the DRC hosted more than a half a million digital items from 20 academic institutions and cultural heritage organizations. Materials include digitized manuscripts, digitized secondary source material, audio, video, official records, documents, and images. The final product of the grant-funded activities was to be a prototype or a set of technical specifications for a new DRC user interface. This new interface, a "Scholar's Dashboard," would allow a user to select entire collections or parts of collections, combine them, add their own descriptions or data, and visualize and present the information in various ways, such as map overlays, timelines, and social networks. Development would be driven by the needs and functional requirements of the participating teams of scholars and collection specialists. The prototyping process was to be dependent on the collaborations among teams: scholars with librarians and archivists focused on collections, and developers and technologists with communities of practice, focused on tools. In addition to this end product, the process was to pair scholars and librarians/archivists and brings together teams with different sorts of subject expertise to work directly with technologists, thus drawing simultaneously on the expertise of end users, collection creators, and technology staff. This was to be a model for future DRC modifications and innovations.

The project's deliverables and performance were affected by two major developments that occurred after the grant was awarded. The first was a significant shift in project personnel. Project Director (PD) John Magill left OhioLINK. Project Manager (PM) Gwen Evans, who at the time of the project's conceptualization was Coordinator of Library Information and Emerging Technologies at Bowling Green State University (BGSU), took a new position with OhioLINK, and soon became Interim Director and then Director of OhioLINK. Accordingly, in summer 2012, Gwen Evans became Program Director. Andy Schocket, Director of American Culture Studies and Associate Professor of History and American Culture Studies at BGSU took over as Project Manager. As indicated in the white paper, this new arrangement came with advantages and disadvantages. On the plus side, having a humanities scholar as PD worked better in the spirit of the project in that it ensured that scholars' concerns would be central to the development of the main deliverable, that is, functional requirements for a dashboard serving scholars' needs. Less positively, having a PD less directly involved with the DRC than Evans had been in her previous position resulted in some disconnection between the project and developments in the DRC.

All other things being equal, slightly less communication between the PD and the DRC would not have been consequential, except for the other major development that affected the project: the dissolution of the DRC. During the spring and fall of 2012, OhioLINK and its participating institutions reevaluated the DRC. The main challenge was one of financial sustainability. OhioLINK had not required participating units to contribute to the DRC's maintenance. It had also assumed that the DRC would benefit from economies of scale, and that after its initiation, further additional collections or expansions in existing collections would only add incremental costs. However, OhioLINK and the DRC found that costs increased in direct proportion to the increase in collections. The platform on which the DRC was mounted, D-Space, required significant maintenance and had limited capacity for easy modification at a time when OhioLINK's finances were under increasing pressure. In short, OhioLINK no longer had the money to support the project, and the member institutions were unwilling or unable to contribute to continuing costs. Accordingly, after a day-long stakeholder summit about the future of the DRC, the Library Advisory Council voted to dissolve the institutional DRC program and OhioLINK on March 8, 2013. Institutions were given a year to migrate their content into alternative platforms. To date, many institutions have chosen to implement multiple platforms depending on the content – platforms in use at institutions include bepress Digital

Commons, open source Omeka, OCLC's ContentDM, open source Fedora/Hydra, and locally supported open source DSpace installations.

The grant's activities begin in spring, 2012, with the conceptualization of how to implement the project. In July and August, Schocket, Evans, and project advisor John Millard held several discussions to consider the structure of the three workshops. After those discussions, as well as consultation with others, it was decided that each two-day workshop would be divided into four sessions. The first morning would consider the humanistic questions that scholars would want to ask of the data. The first afternoon would be dedicated to considering the relationships between those questions, the kinds of materials in the DRC collections relevant to those questions, and their related metadata. The second morning would move to more technical questions of the formats of the digitized (or born-digital) materials in the collections and the technological, analytical, and programming challenges in constructing desirable queries across collections in useful ways. The last session, on the second afternoon, would move to issues of interface: given the questions, the query results, and the format of that data as well as the source material, the workshop participants will consider possible interface features to be included in a scholar's dashboard.

Schocket, Evans, Millard, and others also considered the themes of the three workshops. The first two, as written into the grant, were space and time. Given the considerable analysis already dedicated to spatial and chronological analysis in the digital humanities, these were still considered to be appropriate with little modification. However, the initial theme of the third workshop, "identity," was somewhat reconsidered, because of the concern that it was too nebulous given the kinds of analyses likely to be done on the collections. Accordingly, the theme of that workshop was modified to consider the intersections of identity and social networks, thereby lending the workshop to the kinds of analysis more conducive to computer-assisted research.

Another decision concerned the technologists to be included. The original suggestion was for several software specialists familiar with the structure of the DRC. However, after discussion, the grant team decided that just as important was expertise in programming with the kinds of analysis likely to be included in a scholar's dashboard, such as GPS and TEI. These kinds of expertise will be considered, and the team has started to look for candidates who would be appropriate.

The grant proposal had emphasized that the main deliverable would be technical specifications for an interface and software to be installed on the DRC. However, it also outlined the workshops as "design-and-build," with the object of constructing tools or at least prototypes of tools during the workshops. After further discussion of the project team, and consultation with NEH program officer Perry Collins, the project team opted rather for the workshops to focus on the writing of technical specifications for a tool. The decision to do so was based upon two factors. First was that the programming and design work for a prospective tool or set of tools was beyond the scope of even three two-day workshops. Secondly, even the building of prototypes would require formidable skill and experience with design and programming, and so the participants in these workshops would have to have been people already deeply involved in digital humanities tools. However, the spirit of the grant was to consider not only the needs of those experienced with digital humanities tools but also, and indeed especially, those humanists who would be interested in using digitized and born-digital collections who have little or no such background. The team accordingly decided to focus on technical specifications that could then be implemented dependent upon further funding. The other modification of the deliverables, also discussed with Ms. Collins, was a change in the proposed videos, a change borne both of rethinking and of circumstances. The grant proposal called for three youtube videos, one with film footage from each workshop. On the morning of the first workshop, the videographer cancelled, citing, among other factors, not having gained access to a camera. That in turn prompted reconsideration of the videos, in terms of length and audience. For the second workshop, we did a video capture of the screen with accompanying audio, which we will also do for the third, after which we will compile videos for mounting, with highlights, as originally planned.

To help facilitate these workshops and to write the technical specifications document, the grant contracted with Axia Consulting. Axia has considerable experience facilitating exactly these kinds of exercises, aiding groups to consider information technology needs and solutions, and has done so for various kinds of clients, including commercial, non-profit, and state agencies. This decision has been justified several times over. Despite their initial lack of experience with the needs and resources of academics and libraries, the Axia consultants who have worked on the project (Eric Laus and Shawn Hopper) have more than proved their worth. Their knowledge of information technology systems and how people and organizations interact with them is extensive, and, in addition, their knowledge of the use of information technology in industry brings what had been a missing element in the initial proposal.

The grant was written under the assumption was that the DRC would continue primarily as it had: a central repository to house the digital collections of OhioLINK institutions. However, over the past year, both OhioLINK and the participating institutions have found that model to be unsustainable. OhioLINK had hoped that an increase in the number of collections would result in economies of scale for maintenance and storage. However, costs turned out to be proportional. Furthermore, OhioLINK is under increasing financial pressure and would be unable to increase resources dedicated to the DRC. For their part, participating institutions remained wary of the DRC because of uncertainty over retention policies, difficulties with customization, and concerns over the long-term viability of its underlying platform, D-Space. In January (after most of the planning for the February Scholar's Dashboard workshop had already been completed), the DRC's advisory body opted to pursue a federated model rather than a centralized one. Accordingly, the focus of the workshops was shifted toward technical and functional specifications for software across multiple installations, rather than for one centralized location.

The Scholar's Dashboard has met its objectives as outlined in the grant proposal through the facilitation of the three proposed workshops. The first workshop was held on November 8 and 9, 2012. A total of 23 people attended. The agenda was conceived as a funnel, starting with a very open-sky dreaming session for what scholars would want, and getting more narrowed over the two days to consider particular features. These exercises were highly successful in allowing us to compile a list of desirable functionalities. While many of these were related to spatial analysis – the theme of the first workshop – many were applicable to any kind of use by humanists (for example, allowing users to have accounts with saved searches, and a recommendation engine for suggesting additional sources and objects). Respondents to our post-workshop evaluation survey were for the most part positive about the experience, but some wanted more of a hands-on experience with one or more tools.

Accordingly, for our second workshop, held on February 7-8, we invited Mark Tebeau of Cleveland State University to demonstrate Omeka, for which the OhioLINK programmers had installed a custom instance specifically for Scholars Dashboard participants. The theme of this workshop was "time." Twenty-four people were present for the second workshop. Unlike the first workshop cohort, which included architecture scholars, a cultural geographer, and others primarily interested in space, the humanists for the second workshop were primarily historians. Many of the functionalities that these scholars suggested were similar to those in the first workshop. The participants greatly appreciated the exposure to Omeka which, because of previous work, its extensibility, its many plug-ins, and its potential inter-operability with Fedora, may offer much of what users want. However, many of the features that scholars had suggested in the first two workshops were ones with which they were familiar from commercial offerings, such as Google, Amazon, Pinterest, Historypin, and others.

As a consequence of the findings of the second workshop, planning for the third workshop resulted in conversations with representatives from Google, Amazon, and other commercial firms that offer analytics or interface elements that might be of interest to scholars. Held on May 9-10, the third workshop began with discussions around the workshop's theme, identity and social networks. For the first morning of the session, participants discussed not only current projects relating to these topics, such as the Republic of Letters and the African Slave Voyages Database, but also tools, such as NodeXL. The following sessions featured discussions with representatives of firms that use or adopt such services: Maven Wave, a contractor working with Google Cloud and Google Apps, and Cloudnexa and 8kMiles, contractors working with Amazon Web Services. Although these software-as-services may initially appear more expensive than open-source software, they may not necessarily be most cost-intensive over the long run because they would require less on-site maintenance and expertise. Having a lower up-front cost for mostly pre-designed tools, and having very low rates for data storage and bandwidth, these software-as-a-service possibilities might actually prove more cost-effective for university libraries than developing their own software or having to modify open-source software significantly, continuing software and data retention maintenance costs, and hosting their data on their own equipment. The last session consisted of a discussion with Gwen Evans, concerning the possibilities and difficulties of sustainability given OhioLINK's resources, institutional resources, and the changing funding and technological environment.

**Accomplishments**

Although the DRC has been dismantled, the project was a success in its development of a set of functional and technical requirements for online repositories.

The main deliverable of Scholar's Dashboard, a set of technical requirements for an interface for online collections, was drafted by Eric Laus of Axia based upon notes taken in the workshops and continuing follow-up conversations

with project personnel.  The list of functional and technological requirements was further refined and revised by project personnel.  This report is attached, and the technical requirements have been mounted on the Scholar's Dashboard website, distributed to all participants, and mounted online.

The report includes eleven functional categories: three general, eleven curation, six ingestion, ten metadata, three quality assurance, two reporting, eighteen search, seven security, five socialization, eight user interface, and twelve visualization requirements.  It also includes fifteen technical requirements, for a total of 100 requirements.  Given the difficulty of implementing many of these requirements, their varied nature, and that not every collection has artifacts amenable to the various kinds of searches and visualizations, no single project could meet all of these requirements.  However, one of the additional considerations discussed multiple times was the necessity for modularity and extensibility: that to be robust and relatively future-proof, installations must have the capability for basic functionality with additional functionality added as resources permit.  In addition, among the requirements was support for add-ons and for well documented API's, to allow for a user community also to increase functionality over time.

One notable but perhaps not surprising result of including many scholars savvy about technology but for the most part not digital humanities practitioners, as well as consultants knowledgeable with public and private sector practice, was the desire for functionalities already significantly developed by commercial firms.
They pointed out the difference in maturity and sophistication between technologies developed by and for digital humanists and those developed by commercial firms.  Chief among these were personalized recommendation engines, such as those used by Amazon.com or Netflix, to suggest materials to users based both upon their own history in a repository and those of others who have accessed the same or similar resources in the same depository.  They also admired the kind of sophisticated statistical visualizations available to large-scale corporate and government enterprise systems.  The question then becomes, to what extent can digital humanists develop these kinds of tools with miniscule resources compared to the tens of millions of dollars invested by commercial firms.

Similarly, especially in light of the dissolution of the DRC, whose development, hosting, and maintenance costs had rendered the program unsustainable, participants were interested in private-sector software-as-a-service and hosting models.  Rather than relying on open-source software that often requires modification and the further development for which is often abandoned, humanities scholars and librarians alike saw potential value in contracting with commercial firms to subscribe to software.  Librarians and technologists also cataloged the difficulties of maintaining server space for colleges and universities, especially smaller ones, in terms of cost, expertise, and especially data back-up and retention.  After talking in our third session to contractors who work with Google and Amazon to provide online data services, the entire group was impressed by the capabilities both in terms of data visualizations and data access and retention.  These can be provided at lower overall cost to institutions compared to adapting and maintaining their own repositories.  There are some digital humanities projects that already use such services, like Zotero's reliance on Amazon Web Services for storage, and many institutions that use BePress, a commercial entity, to host their digital institutional repositories.  We may see more use of commercial services for digital humanities support in the future.

As the above passages suggest, the conversations highlighted that among the greatest challenges to the long-term sustainability of digital repositories, especially ones that span multiple institutions, are not technological but rather financial and administrative.  For a centralized model to work, the coordinating institution must have significant resources in terms of personnel (and possibly physical infrastructure).  For a distributed model to work, the various institutions must have a way to share human resources and to coordinate the project.  In either case, the institutions must develop a model for contributing resources either in cash or in kind, while not charging end users directly.  These difficulties have been particularly acute in recent years, with cutbacks to library budgets and human resources at the same time that subscription fees for various commercial databases and journals has increased.  Without significant investment on the part of host institutions, or some other long-term, stable revenue stream, multi-institutional repositories will continue to be extremely difficult to implement and manage over the long term.

**Audiences and Publicity**

The Scholar's Dashboard was publicized in a variety of ways. To attract participants to the workshops, which were limited to practitioners in Ohio, announcements for each workshop were publicized through H-Net. They were also sent through the OhioLINK Digital Resources Management Committee to be distributed to college and university library personnel at OhioLINK institutions, and sent informally to networks of humanities scholars in Ohio, for example, the Hueston Woods group, an informal organization of the history departments of public universities in Ohio. The project established a website ([www.scholarsdashboard.org](www.scholarsdashboard.org)) and was featured in BGSU's Zoom news. The full list of functional and technical requirements generated by the workshop has been distributed to all workshop participants, as well as to OhioDIG, the consortium of Ohio institutions now sharing best practices and resources regarding online repositories, and is posted on the project website. It will also be presented in the poster session at the 2014 Digital Humanities Summer Institute Colloquium.

The project's audience was primarily the librarians and archivists specifically working on the DRC, over fifty professionals at over twenty institutions in Ohio, with the broader audience of current and future users of the DRC as well as all OhioLINK members considering participating in the DRC. At the time, there were 26 participating member institutions, with a waiting list of over a dozen.

**Evaluation**

No formal evaluation activities were built into the grant as originally conceived. However, after each of the first two workshops, participants were surveyed through an online instrument to consider improvements and changes for the subsequent workshop. Respondents to the evaluation of the first workshop noted that workshop objectives were by and large clearly communicated, but urged for more hands-on consideration of particularly technologies and more participation from the OhioLINK technologists. In part, the technologists' reticence was at the suggestion of the Axia consultants, who feared that too much of a focus on specific technological constraints would dampen more broader ranging discussion. It was largely in response to the desire for work with a potentially useful technology to consider its possibilities and limits that Mark Tebeau was invited to demo Omeka for the second workshop. Responses to the second workshop indicated that its activities were much more closely aligned with participants' expectations. For the third workshop, project personnel made sure to have shorter sessions in response to informal feedback from some of the participants of the first two workshops.

**Continuation of the Project**

The DRC's dissolution disrupted original hopes for Scholar's Dashboard to be implemented on a broad scale. However, the results of the project will be used in a variety of ways by the Ohio library and archival community. OhioLINK still hosts some remnants of the DRC, among them a significant collection of e-books, for which it is following some of the technical requirements especially in terms of preservation as a set of principles. In addition, OhioLINK is looking closely at the possibility of serving as a DPLA hub for member institutions, and it is under discussion during the current strategic planning process. Should those discussions lead to implementation, the Scholar's Dashboard specifications will be among the resources providing a template for development.

OhioDIG, the Ohio digitization interest group for archivists, librarians, and others interested in cultural heritage materials that is in some ways the successor to the DRC now that institutions are curating collections locally, will be distributing the Scholar's Dashboard guidelines to its members for their consideration as they implement their own digital repositories. Furthermore, Andy Schocket will be participating in OhioDIG's November, 2014 gathering to discuss user needs and the Scholar's Dashboard's functional requirements for future projects.

**Long Term Impact**

The Scholar's Dashboard has potential long-term impact in several ways. Although the project will not directly lead to additional support or activities directly relating to the workshops or to the DRC, it will have a continued life with both OhioLINK and OhioDIG. For the former, continued activities with its remaining collections for future federated digital repositories will rely on the Scholar's Dashboard technical and functional requirements, as well as the Scholar's Dashboard model of bringing together not only archivists and librarians, but also, crucially, scholars from a range of disciplines to ensure that functionalities reflect what how users would like to interact with OhioLINK repositories. In addition, it provides further insight for OhioLINK as it considers future functionalities for its existing collections or additional installations that it might implement.

Similarly, the Scholar's Dashboard will have a long life with OhioDIG, as a guide for member institutions as they add functionality to current repositories or develop new repositories. Before the DRC's dissolution, liaisons reported robust usage statistics per installation, especially for collections with national or international interest, such as the Wright Brothers Collection a the Wright State University Libraries (now housed locally at CoreScholar, the campus repository Digital Commons bepress platform); the combined viewership of the distributed collections can only grow with time.

**Grant Products**

The main grant product was the list of technical and functional requirements for federated institutional repositories. The list is posted on the project's website, [www.scholarsdashboard.org](www.scholarsdashboard.org), along with a poster presented at the Digital Humanities Summer Institute and Axia's final report. The project will continue to maintain its website, and project personnel will work through OhioLINK and OhioDIG institutions.