The International Journal of Digital Curation
Issue 2, Volume 6 | 2011

Responding to the Call to Curate: Digital Curation in Practice at Penn State University Libraries

Patricia Hswe & Michael J. Furlough,
University Libraries,
Pennsylvania State University

Michael J. Giarlo & Mairéad Martin,
Information Technology Services
Pennsylvania State University

Abstract

This paper describes how the Pennsylvania State (Penn State) University Libraries and the university's central information technology organization, Information Technology Services, are putting into practice key tenets of digital curation through the newly established Content Stewardship program, a joint strategic initiative to implement stewardship services for the university. First, we provide an account of the planning, preparation, and prototyping that informed the initial year of the program. Second, we report on the hiring of a Digital Collections Curator and a Digital Library Architect and how they are advancing the program by putting digital curation into practice, which includes the work of community building. Finally, we address the organizational context of curation in practice, in particular with respect to the challenges of starting and sustaining a stewardship services program for all of Penn State.[1]

[1] This paper is based on the paper given by the authors at the 6th International Digital Curation Conference, December 2010; received December 2010, published July 2011.
The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. ISSN: 1746-8256 The IJDC is published by UKOLN at the University of Bath and is a publication of the Digital Curation Centre.

Introduction

In the last six years, arguably with the inception of the Digital Curation Centre at the University of Edinburgh in 2004, a call to curation has been issued, particularly the curation of data and content in digital formats. Prominent among these calls is the National Science Foundation's solicitation for institutions to partner in establishing "exemplar national and global data research infrastructure organizations" as part of the Data Preservation and Access Network (also known as DataNet).[2,3] Other funding agencies in the United States, such as the Institute for Museum and Library Services and the National Endowment for the Humanities, have responded by facilitating the expansion of library and information science education programs to include curation concentrations (e.g. the Laura Bush 21st Century Librarian Program, Institute for Museum and Library Services), or by offering challenge grants addressing the "data deluge" question and soliciting approaches to the analysis and accessibility of data on a large scale (e.g. the Digging into Data program, Office of Digital Humanities, at National Endowment for the Humanities).[4,5] Consequently, data curator and digital curator positions are on the rise in the library and information science field and have become a topic of research interest (Cragin, Palmer, Varvel, Collie & Dolan, 2009; Lee, 2008). Literature about the practice of data curation (Swan & Brown, 2008), as well as about the implementation of data curation programs in libraries (Gold, 2009; Walters, 2009), is also emerging.

This paper describes how the Pennsylvania State University (Penn State) Libraries (the Libraries) and Penn State's central IT organization, Information Technology Services (ITS), are putting into practice key tenets of digital curation through the newly established Content Stewardship program, a joint strategic initiative to implement stewardship services for the university.[6,7,8] First, we provide an account of the planning, preparation, and prototyping that informed the initial year of the program. Second, we report on the hiring of a Digital Collections Curator and a Digital Library Architect and how they are advancing the program by putting digital curation into practice, which includes the work of community building. Finally, we address the organizational context of curation in practice, in particular with respect to the challenges of starting and sustaining a stewardship services program for all of Penn State.

The Content Stewardship Program at Penn State

The Digital Curation Lifecycle Model presents a cogent framework for carrying out digital curation activities. As the description accompanying it asserts, "The model enables granular functionality to be mapped to it," a mapping that entails defining roles and responsibilities, and building an infrastructure for making operational an organization's agreed-upon standards and technologies.[9] Equally important, if not more so, is the work of laying a programmatic foundation for practicing digital curation within an institutional context – that is, creating the optimum possible environment for following such a model.

[2] Digital Curation Centre: http://www.dcc.ac.uk/.
[3] DataNet: http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503141&org=OCI.
[4] Laura Bush 21st Century Librarian Program: http://www.imls.gov/applicants/grants/21centuryLibrarian.shtm.
[5] Digging into Data: http://www.diggingintodata.org/.
[6] Penn State University Libraries: http://www.libraries.psu.edu/psul/home.html.
[7] ITS: http://its.psu.edu/.
[8] Content Stewardship Program: http://stewardship.psu.edu/.
[9] Digital Curation Lifecycle Model: http://www.dcc.ac.uk/resources/curation-lifecycle-model.

Toward cultivating such an environment, the Libraries and ITS together launched the Content Stewardship program in 2008-2009, initially intending that the program address scientific data management requirements at Penn State. Recognizing that both organizations could apply essential expertise and resources toward meeting those requirements, the Libraries and ITS proposed the program as a joint strategic initiative, a component of their respective five-year strategic plans (Pennsylvania State Information Technology Services, 2008; Pennsylvania State University Libraries, 2008). Because both organizations had already assumed responsibilities for managing a wide range of existing content created as part of pre-existing digitization and instructional programs, it was recognized that a more comprehensive stewardship approach was called for. Thus the program now encompasses stewardship of more traditional digital library collections, such as: digitized special collections; the output of scholarly communications, including journals, monographs, and electronic theses and dissertations; material from the University Archives, such as electronic business records.[10] Building on services and infrastructure already in place, the Content Stewardship program aspires to develop a "cohesive suite of discovery, access, preservation, curation, repository, archival, and storage services for born-digital data" – services to support the entire lifecycle of the digital object.[11]

The goals of the Content Stewardship program broadened for two reasons: the state of existing digital library applications at Penn State, and the desire to implement extensible services and applications going forward. At the outset of the program, our digital library application ecosystem could largely be characterized as a series of siloed applications, all deployed between years 2000 and 2006 and serving disparate needs, with each application requiring substantial resources to manage and maintain. The capacity of these applications to meet new requirements was limited, a fact that was confirmed in a formal platform review conducted in the second year of the program (discussed below). To a significant degree, they supported user-access requirements but were not designed or implemented to support long-term preservation and archival needs. Moreover, Penn State had not implemented institutional repository services around early solutions, a circumstance at once challenging and fortuitous. The challenge lay in the absence of a unified technical architecture and data model and in the absence of digital preservation strategies. The opportunity, however, lay in being able to leverage lessons learned in second or third-generation repository services at other institutions, as well as more mature principles and practices emerging from the digital preservation community.

Our first order of business, therefore, was the design and development of a cohesive information and technical architecture to support our service goals, and we elected to hire a Digital Library Architect to lead this work (this position sits in the Digital Library Technologies [DLT] unit, a part of ITS that primarily supports technical infrastructure services for the Libraries).[12,13] In order to develop a complementary approach to content models and case-based assessment, we simultaneously sought a Digital Collections Curator (a position in the Libraries' Scholarly Communications division).[14,15] The current dearth of such practitioners,

[10] University Archives and Records: http://www.archives.upenn.edu/.
[11] University Libraries' Strategic Plan 2008-2013: http://www.libraries.psu.edu/psul/admin/stratplanjune08.html.
[12] DLT: http://www.dlt.its.psu.edu/.
[13] Digital Library Architect Position: http://www.mail-archive.com/code4lib@listserv.nd.edu/msg04529.html.
[14] Digital Collections Curator Position: http://www.mail-archive.com/code4lib@listserv.nd.edu/msg04689.html.
[15] Scholarly Communications: http://www.libraries.psu.edu/psul/scholar.html.

however, resulted in a year-long search, and neither position was filled until early 2010.

This delay limited progress on programmatic goals, but in the interim we decided to focus on enhancing storage infrastructure and storage management practices. In 2008-2009, Penn State investigated and prototyped the eXtensible Access Method (XAM), a middleware standard from the Storage Networking Industry Association for storage object management.[16] Within the storage industry and in communities of practice, such as the Preservation and Archiving Special Interest Group, more attention has been given to Information Lifecycle Management, which effectively translates to support for digital preservation and archival policies at the storage layer.[17] The XAM prototype project successfully demonstrated that, with the addition of storage management metadata (file type, file size, time of ingest, etc.), objects could be routed to tiered storage devices and could support policy-based retention controls. Yet it also exposed a high barrier to entry with the XAM. Some of this is common to any infant standard or protocol – little adoption, a small community of practice, and immature vendor offerings – but testing the XAM to scale also proved to be challenging. In retrospect, our storage prototyping efforts reinforced the need for an overall technical architecture and for the articulation of higher-level digital preservation strategies to drive storage and archiving decisions, lest storage and archiving decisions drive our digital preservation strategies. This pilot thus laid groundwork for the initial assessment and planning efforts undertaken by the architect and the curator in the program's second year.

Deploying a Digital Collections Curator and a Digital Library Architect

The Work of the Curator and the Architect

In early 2010, the Libraries hired a Digital Collections Curator (co-author Patricia Hswe), and ITS hired a Digital Library Architect (co-author Michael Giarlo). These positions may be seen, at base, as two parts constituting a whole. Hswe is focused mainly on the use and users, or "front-facing," aspects of stewarding digital scholarly content, while Giarlo is largely responsible for the design and development of the technical architecture, or "back-end" aspects, for digital curation applications, including middleware, systems, and services.[18] Below is an account of the work they are accomplishing together; a systematic review of the platforms currently used in the Libraries to deliver digitized content, and the development of a curation microservices test bed. In addition, both Giarlo and Hswe are involved in efforts to foster a community of practice around the technical side of curatorial practice, through informal *unconference* venues (e.g. Curation Technology Camp [CURATEcamp]), drawing participation of digital curation practitioners from a variety of institutions.[19,20]

[16] SNIA XAM Initiative: http://www.snia.org/forums/xam/.
[17] Preservation and Archiving Special Interest Group: http://sun-pasig.ning.com/.
[18] The foregoing is but a précis of the curator and architect roles, providing an overview of the complementarity between their jobs. Besides having these responsibilites, however, each person also has duties that do not involve the other.
[19] An unconference is "a participant-driven meeting": http://en.wikipedia.org/wiki/Unconference.
[20] CURATEcamp: http://groups.google.com/group/digital-curation/web/curation-technology-sig.

The Platform Review

The Libraries currently use four different systems for ingesting and delivering a variety of content and data: CONTENTdm, mainly for image collections, online exhibits, maps, and multimedia materials; Digital Publishing System (DPubS), for journals, monographs, and conference proceedings; Electronic Theses and Dissertations database (ETD-db), for theses and dissertations; Olive ActivePaper Archive, for historic newspapers. CONTENTdm and Olive ActivePaper Archive are proprietary software, while DPubS and ETD-db are open-source platforms.[21]

In early 2010, Giarlo and Hswe conducted a review of these platforms as a first step toward planning for content stewardship services. Digital curation in practice typically calls for assessment of data and content, mainly toward determining possibilities and areas of reuse and repurposing. On a programmatic level, however, assessment of delivery systems is also crucial and arguably constitutes a curatorial measure; for example, such investigations can help ascertain which platforms may be repurposed for delivery and management of other content types. By systematically evaluating these platforms, Giarlo and Hswe were able to understand more fully the needs and expectations of the stakeholders involved, including (but not limited to) end users of these tools, and internal staff who must deliver content via these tools. The work required for the review also helped orient them in their new roles and in the Libraries environment.

Through platform demonstration sessions and interviews with a range of stakeholders, including internal users (such as programmers, database specialists, digitization staff, metadata librarians, public service librarians, subject librarians, and staff at the Graduate School and at Schreyer Honors College) and end users (such as teaching faculty and students), Giarlo and Hswe analyzed the delivery platforms by exploring usage, usability, systems administration, local development, software community, software evolution, interoperability, integration, and organization of information – all of which served as criteria that helped to structure the process of the review. An instrumental resource for these elements was Purdue University's Comparative Analysis of Institutional Repository Software, from which a matrix for analysis of the platforms was derived (Singh, Witt & Salo, 2010). The review also captured user input on common problems experienced across the platforms, as well as potential and current workarounds and fixes.[22]

The evaluation demonstrated that each platform adhered to some, but not all, criteria presented in the matrix. Gaps common to the two most heavily used platforms, CONTENTdm and Olive ActivePaper Archive, had mainly to do with locally unexplored functionalities in each system. A clear next step is to investigate these missing functionalities and enable them, where applicable to workflows, user needs, and management efficiencies. The review also confirmed that two of the platforms, DPubS and ETD-db, are essentially moribund. With these systems, the next step is to test CONTENTdm's potential for e-journal delivery, as DPubS is phased out, and to investigate a replacement for ETD-db and migrate ETD-db content to that replacement.

[21] CONTENTdm: http://www.contentdm.org/; DPubS: http://dpubs.org/; ETD-db: http://scholar.lib.vt.edu/ETD-db/index.shtml; Olive ActivePaper Archive: http://www.olivesoftware.com/products/activepaperarchive.asp.

[22] Comparative analysis of institutional repository software (criteria for analysis are listed on the left-hand side): http://blogs.lib.purdue.edu/rep/.

The Curation Microservices Pilot

The results of the platform review process reinforced the need to investigate alternatives to the siloed platform approach for curation of digital data and content. One promising candidate that has gained momentum in the digital curation community in the last year is *curation microservices* (or *micro-services*), in particular as described and documented by the University of California Curation Center (UC3).[23,24] In a curation microservices environment, a microservice is characterized as a small, independent, interoperable service scoped to encapsulate a single function. A suite of microservices can provide a flexible, non-monolithic framework for coordinating services that carry out the actions reflected in the Digital Curation Lifecycle Model – for example, sequential actions such as ingest, store, and transform, and full lifecycle actions such as description and representation information, which is akin to the characterization microservice in UC3's implementation of curation microservices. A microservices approach to curation, as opposed to a monolithic approach, also reflects a view that rejects a "one size fits all" mentality; since microservices are independent and interoperate with one another, they may be combined to support any number of digital curation workflows (Abrams, Cruse & Kunze, 2009). Preservation of content is thus emergent as a property of a combination of services rather than a property of a particular system (University of California Curation Center/California Digital Library, 2010).

A curation microservices pilot project was launched in fall 2010 to explore a number of strategic aims of the Content Stewardship program, namely, to: define curatorial requirements; build and test a technical architecture; engage software developers and curators at other institutions; treat data in a systematic, cross-platform manner; explore roles and workflows that cross unit boundaries; build a test bed for curation services for electronic records. The experience of determining curatorial requirements is relayed in detail here, both as an example of curation in practice, and to identify a possible need for best-practice efforts.

To define curatorial requirements, the project team concentrated on gathering use cases (besides the curator and the architect, the project team includes a research programmer/analyst and an archivist). This was done to ground the work of subsequent phases in real-world use cases demonstrating the viability – or lack thereof – of the microservices approach to curation at Penn State.

A *use case* is a scenario or situation in which a user – who, for the purpose of the pilot, is a "curator type" internal to the Libraries – works with systems that ingest and make accessible data and content in digital format, and who is trying to do something with those data and content to effect a particular goal. In a curation microservices environment the focus on use cases is critical; use cases drive the development of microspecifications (code) and help translate them into microservices. Colleagues, such as the Metadata Librarian, University Archivist, and Collections Care Supervisor, were consulted about their curatorial activities. In parallel with these consultations, project team members also set about surveying the literature for models of use cases.

[23] Curation Micro-services: http://www.cdlib.org/services/uc3/curation/.
[24] UC3: http://www.cdlib.org/services/uc3/index.html.

The use case as a concept is well defined and well documented in software development literature (Cockburn, 2001; Battista, 2009), which abounds with examples. However, as was quickly discovered, libraries implementing a digital curation approach do not generally share documentation on use cases, whether that approach is an institutional repository instance or a curation microservices setting. Thus, key questions for the project became the following: What would a constructive use case look like in this space? How might such use case modeling benefit the practice of digital curation?

Certainly, there are case studies (also sometimes called "use cases", or user scenarios), more pertinent to the end-user aspect of digital curation, in which the end user typically is a faculty researcher, such as a scientist. An example is the Research Information Network's Case Studies in the Life Sciences project, which surveys how life sciences researchers apply "information sources and services relevant to their research" toward an analysis, evaluation, and management of the information acquired from such sources.[25] These kinds of scenarios, widespread in digital curation literature, are acutely important, for they guide service development as well – albeit for services based not only, or necessarily, on specifications, but also on qualitative and quantitative data about researchers. Moreover, institutions and libraries share these case studies and their findings but tend not to do the same with documentation on use cases (or, perhaps, use cases typically are not captured in such documentation).

One exception may be the survey of use cases provided by the Complex Archive Ingest for Repository Objects project, which detailed various roles and use cases for the Complex Archive Ingest for Repository Objects tool.[26] A similar undertaking – that is, an account of use cases – could benefit the practice of digital curation, particularly in the curation microservices community, and particularly in a context such as that of the Libraries, where there has been no implementation of an institutional repository (e.g. in the form of a DSpace, EPrints, or Fedora installation).[27]

The project team gathered a set of use cases that mapped to various microservices described in the UC3 documentation (see footnote 23). A simple example is provided below (Figure 1). The narrative of this use case is next to a step-by-step version of it, which may be understood as a curatorial workflow – itself a functional translation of the narrative.

[25] Case Studies in the Life Sciences: http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/patterns-information-use-and-exchange-case-studie.
[26] Complex Archive Ingest for Repository Objects: http://cairo.paradigm.ac.uk/projectdocs/index.html.
[27] Dspace: http://www.dspace.org/; EPrints: http://www.eprints.org/; Fedora: http://fedoraproject.org/.

Use-case narrative	Use-case process (workflow)
"The curator has to manage the migration of CDs and DVDs to server storage. In doing so, he tests to verify that the data have transferred properly at the time of transfer; then he periodically needs to test that data in order to ensure they are still viable."	**"CURATOR MIGRATES AN OBJECT" WORKFLOW** 1. Curator receives optical media disk. 2. Curator determines what data files are contained in the disk. 3. Curator determines extent of the files (including checksum). 4. Curator uploads all data files from the disk to the server. 5. Curator receives confirmation that all files have transferred properly (i.e., verifies number of files, checksums, etc.). 6. Curator opens the files to verify readability. 7. Curator verifies readability of all data files.

Figure 1. Example of a use case in the curation microservices test bed pilot at Penn State Libraries.

The activity of collecting use cases became, effectively, translation exercises for the pilot project team. It could also fill a current gap in digital curation practice. Modeling the construction and application of use cases in digital curation practice and sharing such models could cultivate best-practice efforts in the development of curation microservices. It would complement existing documentation about microservices tools, creating a potential for increasing their efficiency as well as contributing to community-building around this curatorial approach. The illustration of use cases can further educate emerging digital curation practitioners, grounding them in a broader understanding of tools, usability, and services. Because a use case by definition also articulates a goal, which may often map to an objective or deliverable in a project, it offers the potential for monitoring progress in that context (Cockburn, 1997). For these reasons, the pilot project has made use case modeling a priority, by documenting it internally as a key component of microservices development in the Libraries.

In addition to these focused efforts on use case development, during the first phase of the project Giarlo worked with the software development team on the technical infrastructure for the pilot, which included procuring a server as a curation environment, analyzing existing microservices code, and making code available on the server (at the writing of this article, this phase was wrapping up). The next phase of the pilot project will include loading sample data from each of the Libraries' four platforms, as well as electronic business records, into the curation environment and testing microservices software by stepping through workflows created from gathered use cases. At the end of the project, we should have an enhanced understanding of whether or not this approach fits most needs at Penn State. If it does, the project will have put the Content Stewardship program significantly down a path toward adopting this approach, by providing relevant use cases, experience deploying microservices, and functional tools.

Building a Digital Curation Community

The growth of the Content Stewardship program is dependent upon a robust community of like-minded researchers and practitioners, both locally and externally. Because positions such as that of the curator and the architect fit into a new niche, the professional community for each is less well established than for professionals in other, more traditional, areas of the Libraries and ITS. Both Giarlo and Hswe reviewed existing professional communities, but perhaps because of the diverse nature of their work, no single professional group in the United States seemed to offer an appropriate fit. As a result, they set about forming such a community. Through an online forum, they invited approximately 30 peer practitioners to share thoughts on the idea of a new community, its niche in a space of many related communities, and its scope. There was wide agreement among this group that a community of practice for digital curators and architects would be useful as a sounding board for peers to run ideas by, as a locus for best practices, and as a forum for sharing successes and failures in digital curation practice.

Building upon the community that had formed earlier around a digital-curation electronic mailing list server called the Google Group on Digital Curation, the group decided to host a two-day event called CURATEcamp, an unconference on digital curation.[28] It took place at the University of California, Berkeley, in August 2010, with curation microservices as its theme, and drew 75 participants, including curators, archivists, programmers, and repository managers. The event was co-sponsored by Penn State's DLT, the division within ITS which employs the architect, and the California Digital Library, home of the UC3.[29] Designed in a loose unconference style, CURATEcamp enabled and promoted active engagement among its participants – an effective community-forming exercise. The format ensured that a wide range of topics was discussed, reflecting the diversity of the practice of digital curation.

Feedback from CURATEcamp was overwhelmingly positive, and work began immediately on follow-up events. Due to the costs and preparation of a stand-alone event, the planning group has thought it wise to organize CURATEcamp as something of a "road show," attaching it in the form of pre-conferences, post-conferences, and breakout sessions to existing conferences. For instance, the 2010 International Digital Curation Conference will have a full-day CURATEcamp workshop focused on the sharing of best practices, as well as on a discussion of tools and technologies, and the 2011 Code4Lib conference will have a full-day CURATEcamp pre-conference on collaborative development of digital curation software tools.[30,31] Penn State remains dedicated to cultivating a community of practice around digital curation and has supported both Giarlo and Hswe in their efforts to maintain and stay engaged with the CURATEcamp community.

[28] Google Group on Digital Curation: http://groups.google.com/group/digital-curation.
[29] California Digital Library: http://www.cdlib.org/.
[30] Digital Curation 101 Lite: http://www.dcc.ac.uk/events/conferences/6th-international-digital-curation-conference/workshops.
[31] Code4Lib 2011 Pre-conferences: http://code4lib.org/conference/2011/schedule#preconf.

Organizational Contexts for Curation in Practice

Developing and deploying the Content Stewardship program has required the input of a diverse set of stakeholders – a diversity likely to grow as the program matures. From its inception as part of the strategic planning process, the program has had the support of both the Vice Provost for Information Technology/Chief Information Officer, and the Dean of University Libraries, and it has been referenced in the overall university Strategic Plan.[32] Ownership at the senior levels must be matched by implementation, however, at all parts of the organization. At Penn State, we have sought to balance the leadership of a few key individual roles with implementation through a widely decentralized organizational structure. This approach brings both dividends and costs.

Shaped in part by the decentralized culture of the Libraries, Giarlo's and Hswe's work at Penn State is also driven by the heterogeneity of the content under stewardship, and by the multi-campus environment of the university itself (which consists of 24 campuses). In addition, within the Libraries, staff from multiple units are engaged in executing well-established parts of the effort, including Digitization & Preservation (reformatting, content ingest), Cataloging and Metadata Services (metadata and data manipulation), Information Technologies (platform/application management; see footnote 7), Scholarly Communications Services (outreach and client relations; see footnote 15), Penn State University Press (publishing), and Reference, Collections & Research (content selection and user engagement).[33,34,35,36] However, these units have been deployed around a digitization program that has been in place for some time, in which content is created almost entirely by the Libraries (similarly, our development of an ETD service in 2000 for Penn State's Graduate School was built upon pre-existing workflows and practices for the management of printed and bound theses.[37] But establishing a program for the curation of user-generated content and data requires new skills, new staff, and, possibly, new management structures.

Managing these transitions and on-boarding new staff, such as the curator and architect positions, have been important programmatic goals of the senior management in the Libraries and ITS. Cross-departmental, even cross-organizational, work is typical of a digital curator's vocation (Madden, 2007) but in practice is challenging to negotiate, especially for an emerging program and staff. Since the curator and architect positions were created out of needs being addressed by a program, not necessarily by a discrete unit in the Libraries, it has been both symbolically and structurally important to have each report to the senior management of the department in which the position is based. For the architect, this is the Senior Director of DLT (co-author Mairéad Martin), within ITS, while the curator reports directly to the Assistant Dean for Scholarly Communications (co-author Michael Furlough), within the Libraries. This organizational placement has signaled the importance of these roles and of the development of the program they support.

[32] University Strategic Plan, "Goal 6: Use Technology to Expand Access and Opportunities": http://strategicplan.psu.edu/technology.
[33] Digitization & Preservation: http://www.libraries.psu.edu/psul/digipres.html.
[34] Cataloging and Metadata Services: http://www.libraries.psu.edu/psul/cataloging.html.
[35] Penn State University Press: http://www.psupress.org/.
[36] Reference, Collections & Research: http://www.libraries.psu.edu/psul/rcr.html.
[37] Penn State ETDs: http://www.etd.psu.edu/.

To broaden involvement in planning and decision-making, a third administrator, the Assistant Dean for Collections and Technical Services in the Libraries, works with the other two in joint sponsorship of a committee called the Content Stewardship Council.[38,39] The Council is critical to the Libraries' and DLT's ability to manage the program and to create the buy-in needed to deploy a new program in a context where roles are already well-established. This Council, consisting of department heads and some staff from DLT and relevant units in the Libraries, provides the forum in which major internal stakeholders can help to shape the program, set priorities, manage its activities, and together participate in its success.

Building on pre-existing programs and services helps to establish organizational continuity, but it comes with some costs. Complex matrix management structures potentially bring significant overhead to the process, and new activities still compete for time and attention with other demands in respective units. Are there other approaches? A distinct Curation Services unit, if well resourced, could move more quickly to make decisions, establish priorities, and develop new services for the university. Rather than distributing new roles and responsibilities to existing units in the Libraries, such a unit could conceivably forge innovative services without the burden of managing legacy projects. However, establishing entirely new budget lines and a significant number of new staff in the currently constrained fiscal environment at Penn State is a hard sell, even for a strategic initiative. Generally, it is easier to raise one-time funds for infrastructure, rather than continuing funds for personnel.

In such a context Furlough and Martin must more realistically build the program through minimal hiring and gradual reallocations. Furthermore, the sustainability of this program will be dependent upon the degree to which it is integrated into the ongoing operations of both the Libraries and ITS, and it is probable that the development of a separate unit would impede that integration. Political and economic situations will vary among universities in the United States, and different models will work better in other contexts. Nonetheless, the organizational complexity and financial challenges faced by Penn State's Content Stewardship program likely will not prove to be a unique experience.

Conclusions and Prospects

Developing university-wide, enterprise-level curation services is a stated goal of the Content Stewardship program (see footnote 8), but it is not yet evident what that may mean. It remains to be seen whether a program, cross-organizational from inception, can most easily address cross-institutional needs. The issue is not merely one of scaling. The question is whether our curation principles and best practices can easily be adopted in decentralized units, for example, or applied in digital asset management at the enterprise level.

As we set a programmatic foundation for putting digital curation into practice in an institutional context, our next steps will involve efforts in technical infrastructure development and in outreach to other campus units. We will be designing, prototyping, and testing a technical architecture that will tie together a suite of curation services. Giarlo will be working closely with Hswe and a team of software developers, systems administrators, and curators, to design an architecture that will fit many possible

[38] Assistant Dean for Collections and Technical Services: http://www.libraries.psu.edu/psul/admin/adtcs.html.
[39] Content Stewardship Council: http://www.libraries.psu.edu/psul/groups/ecsc.html.

curation workflows and use cases, and will scale out to meet performance needs. Rapid prototyping and testing will follow the design phase, bringing into play our gathered use cases as guideposts to ensure a good fit with Penn State's curatorial needs.

Efforts in outreach will include broadening involvement in our initial pilots and engaging other units and researchers on campus. To date, the program and activities described here have been focused almost entirely within the context of the Libraries and its close working relationship with DLT. We began this way in part because the program originated in that working relationship and because the program's mission tracks that of the Libraries. But other ITS units will also have an important future role, including those that manage high-performance computing, educational technology, and emerging technologies.

Going forward, all players currently involved in the Content Stewardship program are being challenged to respond to an increased university-wide focus on data management. The National Science Foundation's May 2010 announcement of future requirements for more formalized data management plans in research proposals has considerably raised the visibility at Penn State, among faculty, research deans, and the Vice President for Research, of the need for curation services.[40,41] Members of these communities are aware of the Content Stewardship program, its goals, and the commitment of the Libraries and ITS to the program. As these discussions have already commenced, we will soon need to develop a variety of new researcher-based use cases.

Preliminary discussions with researchers almost always illuminate new types of problems, creative potential solutions, and the opportunity to develop potential investigative partnerships. Given that time and resources are limited, it will be critical for us to identify the right opportunities and partnerships that will help us test solutions that may be generalized into broad services. Integrating both pre-existing and prospective content management and delivery services into the holistic Content Stewardship framework has provided a solid and familiar basis from which to start. When looking back at work previously undertaken, the challenges and lessons are easier to identify, and in turn new plans are somewhat easier to form. Emerging programs rarely develop linearly, following instead a somewhat meandering and opportunistic path. Ultimately, our success may be marked by the degree to which our goals, and our organization, can remain agile and flexible while still providing credible, valuable services to the rest of the university.

References

Abrams, S., Cruse, P. & Kunze, J. (2009). Preservation is not a place. *International Journal of Digital Curation, 4*(1), 8-21. Retrieved October 23, 2010, from http://www.ijdc.net/index.php/ijdc/article/view/98.

Battista, J.M. (2009). Use cases. Retrieved October 23, 2010, from http://www.johnmichaelbattista.me/software-engineering/use-cases.

[40] Office of the Vice President for Research: http://www.research.psu.edu/.
[41] "Scientists Seeking NSF Funding Will Soon Be Required to Submit Data Management Plans": http://www.nsf.gov/news/news_summ.jsp?cntn_id=116928.

Cockburn, A. (1997). Goals and use cases. *Journal of Object-Oriented Programming, 10*(5), 35-40.

Cockburn, A. (2001). *Writing effective use cases.* Boston: Addison-Wesley.

Cragin, M., Palmer, C., Varvel, V., Collie, A. & Dolan, M. (2009). *Analyzing data curation job descriptions*. Poster session presented at the 5th International Digital Curation Conference, London, U.K. Retrieved April 12, 2010, from http://www.ideals.illinois.edu/bitstream/handle/2142/14544/IDCC_09_FINAL_Cragin_Palmer_Varvel_Collie_Dolan.pdf?sequence=3.

Gold, A. (2009). *Data curation in libraries: short-term developments, long-term prospects*. Paper presented at the American Geophysical Union Fall Meeting. San Francisco, CA. Retrieved April 16, 2010, from http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1027&context=lib_dean.

Lee, C. (2008). *What do job postings indicate about digital curation competencies?* [PowerPoint slides]. Paper presented at the Society of American Archivists Research Forum. San Francisco, CA. Retrieved April 12, 2010, from http://www.archivists.org/publications/proceedings/researchforum/2008/CalLee-SAA-ResearchForum2008.pdf.

[Madden, L. (2007). *Digital curation at the Library of Congress: lessons learned from American memory and the archive ingest and handling test.* Washington, D.C.: Library of Congress. Retrieved April 25, 2010, from http://www.digitalpreservation.gov/library/resources/pubs/docs/digital_curation2007.html.

Pennsylvania State Information Technology Services (2008). *Penn State ITS strategic plan 2008-2013.* Retrieved October 26, 2010, from http://pr.its.psu.edu/.

Pennsylvania State University Libraries (2008). *University Libraries' strategic plan 2008/2009-2012/2013*. Retrieved April 25, 2010, from http://www.libraries.psu.edu/psul/admin/stratplanjune08.html.

Singh, S.K., Witt, M. & Salo, D. (2010). *A comparative analysis of institutional software.* Retrieved October 17, 2010, from http://or2010.fecyt.es/Resources/documentos/GSabstracts/AComparativeAnalysisInstitutionalRepositorySoftware.pdf.

Swan, A. & Brown, S. (2008). *The skills, role and career of data scientists and curators: an assessment of current practice and future needs*. Report to the JISC. Truro, UK: Key Perspectives Ltd. Retrieved April 12, 2010, from http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dataskillscareersfinalreport.pdf.

University of California Curation Center/California Digital Library (2010). *Merritt: an emergent micro-services approach to digital curation infrastructure.* Retrieved May 23, 2011, from https://confluence.ucop.edu/download/attachments/13860983/Merritt-latest.pdf.

Walters, T. O. (2009). Data curation program development in U.S. universities: the Georgia institute of technology example. *International Journal of Digital Curation 4*(3), 83-92. Retrieved April 15, 2010, from http://www.ijdc.net/index.php/ijdc/article/viewFile/136/153.