Collecting Projects: Bridging the Project/Service Divide

Leslie Barnes
Information Technology Services Department, University of Toronto, Toronto, Canada.

Rachel Di Cresce
Information Technology Services Department, University of Toronto, Toronto, Canada.

Abstract:

Can we convert digital projects into digital collections? Functional websites are always ephemeral and, increasingly, those working in digital scholarship are recognizing the importance of sunsetting project. Can we create a service pipeline in which decommissioned projects are converted into usable and accessible datasets?

Like many libraries, University of Toronto Libraries (UTL) began developing digital scholarship projects in the 1990s and early 2000s by creating digital collections or databases that allowed users to search structured data. There are a number of strategies libraries and digital humanists can use to preserve their work: web archiving, migration, or some combination thereof. At UTL, we have begun to view the data, rather than the platforms that house and/or present them, as the essential product of scholarship. This perspective has led us to prioritize data curation--even outside of presentation platforms. Understanding project lifecycles in terms of data curation allows us to bridge between the project/service divide. In "Supporting Digital Scholarship in Research Libraries: Scalability and Sustainability," Vinopal and McCormick distinguish between specialized projects and scalable, broad-based services (2013). At UTL, we must balance both approaches and thereby ensure the sustainability of digital scholarship at UTL.

We are currently working on strategies in which project data--whether images or other digital objects, or structured or semi-structured data--become objects to be collected. Consequently, we can sunset projects knowing that their data will remain available in the long-term and researchers will continue to comply with funder requirements. We will present our strategies for creating this pipeline and therefore cultivating sustainability in digital projects at UTL.

Keywords: Digital Humanities, Data curation, Sustainability, Digital projects, Library service

At University of Toronto Libraries (UTL), we have a long history of supporting digital projects in a robust local infrastructure. Our earliest digital projects date back to the 1990s and predate the wide acceptance of "digital humanities" as a field, though it was practiced

under different guises. These projects included databases, collections, online exhibits, and websites that presented data, digital media, and other scholarly outputs in both structured and semi-structured ways. In the 2010s, however, two decades of accumulated projects--many of which continued to draw large numbers of users--presented library staff, particularly those in Information Technology Services (ITS), with challenge. As individuals' portfolios grew to accommodate both legacy projects and new initiatives, staff found it increasingly challenging to maintain all of these projects with equal vigor. Prior to our (Leslie Barnes' and Rachel DiCresce') arrivals at UTL (Summer 2014 and Winter 2015, respectively), several earlier projects, notably Representative Poetry Online, Records of Early English Drama, and Documents of Early English Dataset, had been or were in the process of being migrated to new platforms in grant-funded migration projects. Others were left on silent running on older but still serviceable platforms. ITS was simultaneously working on supporting and extending services, including our institutional repository, TSpace, and our online collections service, Collections UofT. Additionally, to enable librarians and researchers to showcase their work on the Web, we began Exhibits UofT, a lightly mediated online exhibit service based on Omeka.

Inspired by Jennifer Vinopal and Monica McCormick's model of tiers of service that they outline in "Supporting Digital Scholarship in Research Libraries: Scalability and Sustainability," ITS became increasingly invested in distinguishing between services and projects as part of a concerted effort to ensure scalable, ongoing services for a large institution with growing user demands for digital project support. Services are ongoing and support user functions in a sustained way. In contrast, projects are characterized by having an identifiable conclusion--whether that is a concrete output or outcome. There is, of course, some gray area between these two categories, but our focus for services was on "tool standardization" (6), which Vinopal and McCormick identify as the key to sustainability. This means minimizing customization, emphasizing reusability, and creating replicable workflows (6-7). In contrast, projects often require hands-on customization, usually funded through grants, and bespoke platforms or solutions. While many in the DH community have described how projects may resist completion (Brown "Published Yet Never Done"), in the case of the projects--both legacy and incoming--that we're describing "doneness" is often linked to very practical concerns: grant funding has concluded or deliverables outlined in a project charter or grant proposal have been generated. What often isn't accounted for is what happens after these milestones have been crossed--how do websites, interactive databases, and tools live on after completion.

Over the last 5 years, a number of events both prompted and enabled us to reevaluate how we deal with projects at end-of-life. In 2015, ITS and the Centre for Medieval Studies at University of Toronto received an Andrew W. Mellon grant in support of Digital Tools for Manuscript Study (DTMS), for which the project team developed a IIIF plugin for Omeka and collation tool, with the aim of enabling researchers and institutions to share manuscript images, annotations, and other IIIF metadata with a strong focus on building up data curation infrastructure. Our Digital Preservation Librarian began Project Canopus, to build a digital asset management system and extend digital preservation infrastructure for UTL. And lastly, several of those "silently running" projects were running on ColdFusion servers, which presented an urgent security risk in 2016. Suddenly, the department was engaged in a flurry of migrations, preservation tactics, and decommissions. In short--ITS was (and has always been) a busy place.

These intersecting circumstances and events brought some known but not necessarily heeded truths home: we needed a strategy for ensuring a balance between services and projects. One of the hallmarks of that balance would need to be ensuring an afterlife for projects. Instead of insisting on a strict division between services and projects, we started thinking about ways of actively crossing that divide and bring projects into the fold of services. Through DTMS, we began to see how powerful viewing data, rather than the comparatively fragile platforms that house and present them, as an essential product of digital research projects and digital scholarship. Understanding project lifecycles in terms of data curation allows us to bridge between the project/service divide and we are now creating project workflows and pipelines that focus on data curation that help enact this intersection.

We are certainly not the first organization to grapple with how to sunset or preserve digital projects. Indeed, this has been a pressing issue among digital librarians, Digital Humanities (DH) scholars and practitioners, and digital archivists for decades. In their recent series of "Archiving DH" blog posts for the University of Virginia Scholars' Lab, Ammon Shepherd, Amanda Visconti, Brandon Butler, and Lauren Work summarize the multi-faceted challenges of keeping DH projects alive for the long term. They point out the array of obstacles to sustaining of DH projects, including the constant changes in technology; often unstable funding sources; and shifting human resources and obscured human labour, which in turn leads to uncertain ownership. All of these problems persist, they point out, despite sustained interest in the preservation of DH projects. And, indeed "keeping software secure and up-to-date is a continual, long-term project. The opposite of 'archived' and 'preserved'" (*Archiving DH Part 2: The Problem in Detail*).This blog post chimes with writings by several librarians and scholars working in DH and digital scholarship. In its 2003 report of findings from a commissioned survey of digital heritage initiatives, the Council for Library and Information Resources warned that funding, institutional support, and technological changes posed serious sustainability issues for digital projects (Zorich, 22-24). Bethany Nowviskie has urged the DH community to focus on forward migrations, maintenance, and metadata improvements as they integral to ongoing DH success rather than solely pursuing the next new thing (Nowviskie). Andy Shocket want scholars to understand the ephemerality of much DH work, while also pursuing sustainability (47).

Librarians and researchers *know* that we need to begin with the foreknowledge that technologies will shift and projects may die. Geoffrey Rockwell insisted that "Projects should be designed from the beginning to die gracefully, leaving as a legacy the research data developed in a form usable in the future" (31). One may be tempted to think that libraries have somehow solved these issues because they are the perceived custodians of scholarly outputs. Operational and technical issues complicate matters, however. Projects developed in conjunction with project partners or non-permanent staff and technology staff are frequently ended without a clear governance plan that integrates the project and its long term home (Craft, 67-68). Ownership and responsibility are often more complicated than first appears if not adequately planned for.

As Leslie Johnston discusses in her 2013 blog post "Digital Humanities and Digital Preservation," the long-term fate of many digital humanities project belongs to one of two often unappealing options: "Preserve the content and the look and feel exactly as they were implemented. This is often close to impossible. Preserve the content but forgo the look and feel. This is often extremely unpopular." Confronted with the unsavory binary of constant migration, which could easily consume entire, even robust, IT departments' time and resources, or the "killing and pinning it" model, which may minimize ongoing costs and

resource dependency, but also minimize usability and consequently also accessibility or meaningfulness, libraries are often in a double bind. Is there a way, we wondered, to bridge between these unappealing (or, in some cases, impossible) alternatives?

I want to circle back to our ColdFusion security crisis--which led to some admittedly pretty *ad hoc* measures and revealed some organizational gaps. While ITS had been thinking more concertedly about projects and their interaction with services, we hadn't given as much attention to projects' afterlives. Some of our solutions have been practice-based and indeed follow the advice of many other librarians and DH practitioners: having concrete inventories, assigning responsibility for project, and ensuring that projects must, indeed, be projects, rather than users' needs being fulfilled by service offerings. At the same time, we are aware that grant-funded projects will arise and that UTL requires a holistic approach to solving this aspect of the sustainability problem. The technological backbone of this approach emerged from three projects which have striven to make infrastructure a service, prioritize project data as an output, and anticipate that interfaces will inevitably become outmoded and platforms will eventually become unsupportable.

The first of these is Project Canopus, a project led by our Digital Preservation Librarian, to build a digital asset management system at UTL that will preserve born digital, digitized assets, and research data in the long term on a tape system. Project Canopus, then, is about building infrastructure as a service and in support of a service. The second and third of these are two Mellon grant-funded projects: Digital Tools for Manuscript Study (DTMS) and The Book and the Silk Roads: Phase 1 (BSR). In both of these projects, the project team has two goals: build technological infrastructure that will reliably transmit, store, and make meaningful humanistic research data of numerous kinds and, perhaps more fundamentally, find a way of both sensitively meeting researchers' immediate scholarly needs and aims while also supplying them with an infrastructure that is both immediately useful and enables their work to endure. We took to heart Miriam Posner's argument in her 2016 talk, "Data Trouble: Why Humanists Have Problems with Datavis, and Why Anyone Should Care", that humanities scholars are charged with making ontological interventions in their disciplines and fields. *Data*, as a concept, is at first glance at odds with this objective. Through both of these Mellon projects, which are as much community-building endeavours as technological development projects, we are trying to find ways to cross that breach. Our ultimate goal is to make infrastructural systems that enable us to effectively make a collections service around our projects--keep meaningful data meaningful and usable, while also allowing the library to provide sustainable service.

DTMS began in 2015 and was always a partnership between ITS and the Centre for Medieval Studies. Our aim was to develop tools that would usefully aid researchers in manuscript studies, supporting digital and semi-digital research workflows. The tools we built were a IIIF Omeka plugins that enabled researchers to access, share, and annotate manuscript images from around the world and a collation tool, VisColl, which allows researchers to input, visualize, export formatted data about the construction of books. Both of these tools were developed around the International Image Interoperability Framework (IIIF), which we also incorporated into UTL's infrastructure. IIIF is a set of APIs that provide uniform access to the world's images and integrate viewing and annotation tools. As a central specification that enables this connectivity, IIIF supports a growing community scholars, technologists and institutions across the world that are engaged in image-based research. ITS also developed its own API, which uses the IIIF specification, but also includes annotation data, and adds metadata pertinent to our local digital collection infrastructure. In this way, we could have

full control over the creation, distribution and preservation of data produce by the project and our other platform tools.

From the outset, the technology team was concerned by the inevitable degradation of technology platforms and other containers, and thus emphasized the longevity and centrality of the scholarly data. Focusing on the data and, crucially, the metadata that describes it, ensures these outputs could be preserved, accessed and reused in the long-term. In our proposal--and in practice--we tried to shift the focus away from the tool as the artifact or output of research and emphasize that the true value of this work lay in the data researchers were creating, sharing, and commenting on. In development, we followed the core tenets of user-focused design, modular, low-weight technologies, the importance of data and metadata standards to ensure interoperability. Though our outputs were tools, we were attempting to foreground the infrastructures and cultures that support and sustain data curation. Consequently, a large part of DTMS focused on community-building. And indeed, the premises underpinning the technologies we built--user-focused design, data interoperability, and modularity--are themselves community-focused principles that allow technologists and users to share code, data, insights, and scholarly information.

ITS is now embarking on another Mellon project in collaboration with CMS, the Book and the Silk Roads (BSR), which will bring together scholars from across areas of study to reframe the story of how the book developed by focusing on binding techniques from across the premodern world. Using non-destructive scientific imaging and analysis techniques to study book bindings from 500 - 1500 C.E, researchers will generate data. The technical team will develop presentation prototypes that will visualize and disseminate this data. Technically, this project engages with large amounts of complex, heterogeneous data that need to be accessed, presented, and stored. The core technical focus, however, is on building robust digital curation infrastructure and ensuring that data is described using metadata standards or specifications that enable cross-platform usability, since data and metadata can endure well beyond the platforms in which they are created.

Although this first phase has tangible development outputs, its primary goal is to fully understand a community of scholars and their needs in addition to the technical challenges supporting this kind of scholarly work will present. We hope that by taking this approach, we will be able to deliver meaningful support at the project level, while also extending our operational and technical infrastructure to ensure that scholarly outputs are meaningful in the long term.

Both DTMS and BSR are, at their core, community-building projects as much as they are infrastructure-building projects. And in both cases they've afforded ITS the opportunity to not only extend services, but do so with a careful eye to enacting a rapprochement between projects and services. To do so, we've discovered, means sensitively engaging with the core reasons why the project model is so appealing--novelty, ontological as well as technological innovation--while not losing sight of the perhaps less glamorous, but no less important, requirements of services: endurance, reliability, and replicability. To do so, we've proposed foregrounding data and creating an infrastructure through which to collect that data *as a service*, while still being able to support project-level work.

learn from. In addition, we would like to thank the developers, system administrators, librarians, information architects and support staff who've been integral in the growth and maintenance of the department and its digital assets.

References

Brown, Susan, Clements, Patricia, Grundy, Isobel, Ruecker, Stan, Antoniuk, Jeffery, and Balazs, Sharon. "Published Yet Never Done: The Tension Between Projection and Completion in Digital Humanities Research". *Digital Humanities Quarterly.* vol. 3 no. 2, 2009.
http://www.digitalhumanities.org/dhq/vol/3/2/000040/000040.html

Butler, Brandon, Shepherd, Ammon, Visconti, Amanda, and Work, Lauren. "Archiving DH Part 2: The Problem in Detail". *Scholar's Lab,* University of Virginia Library, Published March 11, 2019. https://scholarslab.lib.virginia.edu/blog/archiving-dh-part-2-the-problem-in-detail/. Accessed on May 27, 2019.

Butler, Brandon, Visconti, Amanda, and Shepherd, Ammon. "Archiving DH Part 3: The Long View". *Scholar's Lab,* University of Virginia Library, Published April 30, 2019. https://scholarslab.lib.virginia.edu/blog/archiving-dh-part-3-the-long-view/.

Craft, Anna R. (, Contributor) "Digital Scholarship Planning: A Perspective on the CNI-ARL Workshop". *Serials Review*, vol. 44, no. 1, 64-70, 2018. DOI: https://doi.org/10.1080/00987913.2018.1433930

Hrynaszkiewicz, Iain. "Data Descriptors: Providing the Necessary Information to Make Data Open, Discoverable and Reusable." Impact of Social Sciences, LSE Impact Blog, 22 October 2014. http://blogs.lse.ac.uk/impactofsocialsciences/2014/10/22/data-descriptors-open-discoverable-nature/

Howard, Jennifer. "Born Digital, Projects need Attention to Survive". The Chronicle of Higher Education, January 6, 2014. https://www.chronicle.com/article/Born-Digital-Projects-Need/143799

Johnston, Leslie. "Digital Humanities and Digital Preservation". *The Signal,* Library of Congress. *Published on April 12, 2013.* https://blogs.loc.gov/thesignal/2013/04/digital-humanities-and-digital-preservation/

Kirby, Jasmine S. "How NOT to create a digital scholarship media platform: the history of the Sophie 2.0 project". *IASSIST Quarterly*, vol. 42, no. 4, 1-16, 2018. https://doi.org/10.29173/iq926

Kirk, Anne M., et al. "Building Capacity for Digital Humanities: A Framework for Institutional Planning". *ECAR working group paper.* Louisville, CO: ECAR, May 31, 2017. https://er.educause.edu/~/media/files/library/2017/5/ewg1702.pdf?la=en

Moraon, Nancy L and Pickle, Sarah. "Sustaining the Digital Humanities. Host Institution Support beyond the Start-Up Phase". *Ithaka S+R.* June 18, 2014. https://sr.ithaka.org/wp-content/uploads/2015/08/SR_Supporting_Digital_Humanities_20140618f.pdf

Nowviskie, Bethany. "Digital Humanities in the Anthropocene". *Blog, Bethany Nowviskie.*July 10, 2014. http://nowviskie.org/2014/anthropocene/

Nowviskie, Bethany and Dot Porter. "The Graceful Degradation Survey: Managing Digital Humanities Projects Through Times of Transition and Decline'." *Digital Humanities (*2010): 7-10. http://dh2010.cch.kcl.ac.uk/academic-programme/abstracts/papers/html/ab-722.html

Posner, Miriam. "No Half Measures: Overcoming Common Challenges to Doing Digital Humanities in the Library". *Journal of Library Administration,* vol. 53, no. 1, 2013. https://escholarship.org/uc/item/6q2625np

Posner, Miriam. "Data Trouble: Why Humanists Have Problems with Datavis and Why Anyone Should Care". *YouTube,* uploaded by USFDVSS, 19 April 2016. https://www.youtube.com/watch?v=sW0u1pNQNxc

Rockwell, Geoffrey and Day, Shawn. 'Burying Dead Projects: Depositing the Globalization Compendium'. *Digital Humanities,* vol. 8, no. 2. College Park, Maryland (2009), http://www.digitalhumanities.org/dhq/vol/8/2/000179/000179.html

Schocket, Andy. "Embracing ephemerality in the digital humanities". *History, CLASS,* Published February 2, 2016. http://www.andyschocket.net/2016/02/02/20160202embracing-ephemerality-in-the-digital-humanitiesembracing-ephemerality-in-the-digital-humanities/

Senseney, Megan et. al. "Unanticipated Afterlives: Resurrecting Dead Projects and Research Data for Pedagogical Use". *DH 2018*: 122-124. https://dblp.org/rec/conf/dihu/SenseneyMPTW18

Shepherd, Ammon. "Archiving DH Part 1 - The Problem". *Scholar's Lab,* University of Virginia Library, Published January 31, 2019. https://scholarslab.lib.virginia.edu/blog/archiving-dh-part-one/.

Sigrid Anderson and Gomis, Melissa. "So what are you going to do with that?: The promises and pitfalls of massive data sets". *College & Undergraduate Libraries*, Published on July 20, 2017. https://deepblue.lib.umich.edu/bitstream/handle/2027.42/139596/So%20what%20are%20you%20going%20to%20do%20with%20that%20The%20promises%20and%20pitfalls%20of%20massive%20data%20sets.pdf?sequence=1&isAllowed=y

Sula, Chris Alen. "Digital Humanities and Libraries: A Conceptual Model". *Journal of Library Administration,* vol. 53, 10-26, 2013. http://chrisalensula.org/wp-content/uploads/sula-digital-humanities-and-libraries-a-conceptual-model.pdf

Vinopal, Vinopal and McCormick, Monica. "Supporting Digital Scholarship in Research Libraries: Scalability and Sustainability". *Journal of Library Administration*, vol. 53 no. 1, 27-42. (2013).

Zhang, Ying, Lui, Shu and Mathews, Emilee. "Convergence of Digital Humanities and Digital Libraries". *Library Management,* vol. 36, no.3, 362-377, 2015. https://www.researchgate.net/profile/Shu_Liu/publication/277940906_Convergence_of_digit

al_humanities_and_digital_libraries/links/56b2602508aed7ba3fedd01e/Convergence-of-digital-humanities-and-digital-libraries.pdf

Zorich, Diane M. "A Survey of Digital Cultural Heritage Initiatives and Their Sustainability Concerns. Managing Economic Challenges." *Council on Library and Information Resources,* Washington, DC, 2003. https://files.eric.ed.gov/fulltext/ED482236.pdf

Zundert, J. v. "If you build it, will we come? Large scale digital infrastructures as a dead end for Digital Humanities". *Historical Social Research*, 37(3), 165-186, 2012. https://doi.org/10.12759/hsr.37.2012.3.165-186