1-10-2015

Taking Care of Digital Efforts: A Multiplanar View of Project Afterlives

Robin Camille Davis
CUNY John Jay College

Recommended Citation
Taking Care of Digital Efforts
A Multiplanar View of Project Afterlives

Robin Camille Davis, Library, John Jay College of Criminal Justice (CUNY)
robdavis@jjay.cuny.edu

Presented at the MLA Convention in Vancouver on January 10, 2015, as part of the Hacking the Renaissance panel (session #420).

This is the script of my presentation. Illustrative images and charts from my slides are also included. See archived tweets from audience members on Storify (PDF).

—

Introduction

My talk, in your programs, is titled “Curating Digital Efforts in the Library.” But Lake Superior State University has banned the word ‘curate’ in its 2015 Annual List of Banished Words. I’m a librarian, and don’t usually like when things are ‘banned,’ but maybe in this case I might agree. Sometimes we do misuse or overuse the word curate to the point where its meaning is vague, and I know I was misusing the word here. So instead maybe we can call this presentation “Taking Care of Digital Efforts.” And I’ll add a subtitle, “A Multiplanar View of Project Afterlives.”

I took part in Early Modern Digital Agendas in 2013, where I met some of my co-panelists, and the conversation we had there touched on many topics. One topic that really caught my attention was the preservation and transformation of digital projects, and this is the topic I’m following up on here.

We know a lot about the afterlives of books. We’ve got quite a few experts in printing history in this room. We’ve got well-established methods of determining the provenance of a manuscript or a book. After all, books and manuscripts can stick around for a very long time, changing owners, crossing borders.

But the worldwide web is just over 20 years old. We just don’t know what the long-term afterlife of a digital project is yet. We have indications, certainly. So much of the early web is lost to us — as Megan Sapnar Ankerson puts it:

“It is far easier to find an example of a film from 1924 than a website from 1994.”

What’s at stake here? Many of us in this room are dedicated to considering digital projects to be a form of scholarship and scholarly communication, so that our non-traditional online efforts can be considered alongside books and articles when it counts. Books and articles tend to stick around and form a record of a scholar’s work. Digital work can be an amazing and accessible part of a scholar’s path — or if it disappears, it could be a hole in the record. So for those of you who do digital work, examining the afterlives of digital efforts is something you’re already heavily invested in.

---

Interlude

Let’s take a second and think about another possible world. Imagine that you’ve got a digital humanities project on your laptop. You’re building a collection of texts and images. It’s beautifully designed and extremely well-researched. Imagine that the more images and information you put into your project, the heavier your laptop gets. You start to feel the physical force of your work. The more that’s in your laptop, the more it weighs. You have to reinforce your desk at work because you’re working with the weight of 1, 2, 300 books at once. Those digital humanists who work with, say, 25,000 texts have to start drinking protein shakes and joining Crossfit gyms just to continue their scholarship.

Now imagine that you’re done with this project and you don’t want to keep it around — this hundred-pound laptop is a beast to lug around. So you decide to move your project elsewhere. You have to push that weight into another receptacle. It takes all afternoon; you have to ask your new Crossfit friends to help you out; after you’ve cleared out your laptop, you’re all sweaty and gross, but happy to do it all over again to pursue a new line of scholarly inquiry.

This is a silly image. It’s silly because we don’t have this problem. Gravity does not act much at all on our digital projects. Other forces do, but the digital is smooth and weightless and effortless. How easy it is to delete files. Much less effort than putting together a box of free books for giveaway and taking it down to the curb. In fact, you don’t even have to delete digital work for it to be deleted, given time.

---

Digital decay

The ease of digital decay is something librarians have felt disturbed about for a long time. As Neal Beagrie puts it:

“In the right conditions papyrus or paper can survive by accident or through benign neglect for centuries, or in the case of the Dead Sea Scrolls, for thousands of years .. In contrast, digital information will not survive and remain accessible by accident: it requires ongoing active management from as early in the life-cycle as possible.”

Or as Jerome McDonough puts it, a little more forcefully:

“Everything digital dies!”

[McDonough, J. Class lecture.]

And it often dies sooner than we’d like. This is an old but ongoing problem, with ongoing research, and it haunts just about everything I do. All of the work I do is digital — probably yours too. What kinds of forces could preserve the work we do?

I turned to a sample set of projects representative of the kind of work I’m interested in to see how old projects fared. Did they get finished, are they ongoing, or were they abandoned? Are these projects still accessible online?

—

Investigating DH2005 projects

So before I show what I found in my inquiry, let me ask. Was anyone in this room at the 2005 Digital Humanities Conference at the University of Victoria? This was the annual ACH/ALLC Conference. [Edit: one audience member raised her hand!]

Let’s set the scene. The Blackwell Companion to the Digital Humanities was published just the year before. The Virtual Humanities Lab at Brown had also just begun. This was the first year that CLIR offered post-doc fellowships in DH work. Digital humanities, aka humanities computing, was well-established by 2005.

On the other hand, centerNet wasn’t around, and the Scholars’ Lab at UVA didn’t exist yet. This is back when social media was just starting out, and Facebook didn’t have a news feed, and Gmail was invite-only. It has been a while, and DH has changed and grown massively in the last ten years.

So this is the sample set I chose for a preliminary investigation: all the projects presented at DH2005. By projects, I mean something interactive or reusable — in other words, not a paper. You’re going to argue with me that a paper is interactive and reusable — true enough. But here I mean projects like image collections, or text corpora, or software packages, or web applications.

—

Methodology

I went through each project from the point of view of an interested user trying to track it down, just like an interested reader following up on a footnote. Using the abstracts from the DH2005 program, and
aided by Google and the academic databases my library subscribes to, I took on the role of detective and tried to find and assess all the projects. Some were really easy: the link in the abstract still resolved, and bam, I was in. Others were tougher: URLs changed, or the project itself had morphed into something else, but it was still findable. And some just disappeared into the ether.

I’ll note a caveat: I considered all projects to be equal, big and small, though I did not consider projects presented at poster sessions. Some were small research projects presented by PhD candidates as part of their dissertation research. Some were big, grant-funded projects managed by DH rock stars. All of these online projects were hosted by institutions. And last caveat, this is not a shame session for projects that are no longer available. More on that later.

---

**Status & accessibility of DH2005 projects, 10 years later**

All of the projects I considered were accessible online in 2005. How many could I find now?

**Accessibility of online project in 2015**

- Online (paid), 1, 2%
- Not online, 27, 45%
- Online, 32, 53%

Just over half of the 60 projects presented at DH2005 are still online.

One of the projects that is no longer online morphed into a different project, from Nora to the Monk Project. Another project that is not online is ongoing and forthcoming. Only one of the projects still online is paid access only; all the rest are freely available.
Many projects had changed URLs, often without a redirect. This chart accounts for the new URLs I tracked down.

Accounting only for original URLs, accessibility would be much lower. The half-life of URLs is of high interest to web historians and preservationists. A number of studies have attempted to determine the time required for half of all online citations in a given academic journal to disintegrate. It differs by discipline: studies find that half of a law journal’s online citations disappear after 1.4 years; in history, the half-life is 7 years; in medical science, the half-life is 13 years; digital library objects, 24.5 years. [Studies quoted in Sampath Kumar, B.T., and Manoj Kumar, K.S. "Decay and Half-life Period of Online Citations Cited in Open Access Journals." International Information and Library Review 44.4 (2012): 202-11.]

So to use this approach, for the projects listed in the DH2005 program, perhaps we can say that the half-life of those projects is around ten years, as I found that only half are still accessible. What a shame. We can’t negotiate sites of memory if the sites are no longer there.

As of now, what are the statuses of these projects?

One third is unknown, that is, I couldn’t find information that gave me any indication that it was finished or ongoing or abandoned. Two were clearly abandoned.

One third were clearly completed. They finished building out the project and moved on. These projects include Mayors and Sheriffs of London, the Carte Calendar Project, and the Clotel Electronic Edition. These are projects that were finished, in 2005 or later, and are now frozen in time.
And, to my great surprise, almost one third are ongoing! These projects include the Walt Whitman Archive, the Early Modern Literary Studies journal, the Tibet Oral History Archive, and Great Unsolved Mysteries in Canadian History. Maybe this is no great surprise — some of these projects were of an infrastructural nature, made to be built and added to, and they were never meant to be finished.

![Finished projects: accessibility in 2015](image)

Of the 23 projects that were considered to be complete, three-quarters are still online; one quarter is no longer accessible.

—

### Survival factors

There are a number of factors that contribute to a digital project’s survival. Some are technical or technological — a project encoded in Hypercard is much harder to access now, because that software is outdated. But most of what contributes to a project’s decline and fall is social or institutional.

The role of project management in a project’s survival isn’t something I looked at here, but it has been followed up on before. The “Graceful Degradation: Managing Digital Projects in Times of Transition and Decline” survey was sent out in 2009 by Bethany Nowviskie and Dot Porter, with over 100 responses. The survey asked about digital projects in or related to the humanities, and the authors analyzed the findings to see how digital projects fared when facing difficult times (like funding troubles) and periods of transition (like colleagues leaving the institution). These projects weren’t limited to a certain time period.
64% of respondents had experienced the decline of a project or had weathered a period of difficult transition. Of these, 51% were identified as still ongoing with 26% abandoned, 15% finished, and 8% just getting started.

The work: tracking down DH2005 projects

The social ties of CVs: Again, I was using my best librarian detective skills to track down these projects, without actually talking to a human, unlike that survey. In my investigations, for projects that were not available online or whose status I could not determine, a good deal of information about a project was found in the CVs of people who had worked on it.

The social and professional practice of writing CVs is geared toward employment, but it turns out to be an important record for digital projects. It’s a record that is cumulative and historical. It is updated regularly, and for most scholars who do digital work, it’s online. Moreover, nearly all of the scholars who presented at DH2005 are easily findable at their current institutions.

Using a CV as a kind of social tie to a digital work gave me an unexpected amount of metadata. It gave me dates, sometimes a description, and it gave me a trail from 2005 to the present of what happened on that path of inquiry. Sometimes the afterlife of a project is the birth of another project.

Traditionally published papers: For projects that consist of building a data set or a corpus or a collection, it can seem like writing up a paper about it is jumping through a hoop — often because that’s all a tenure committee will consider. The paper can be a bit of an afterthought to the work. Not always — and by no means am I bashing publishing here — but for these digital projects, the meat of the work was the online tool or website or software. Article-writing usually took a back seat to the programming and coding.

However, in many cases, the only evidence I could find that a project ever happened was by finding a paper written by team members. For better or for worse, the traditional publishing process is extremely old and (compared to the web) stable. There are agreements and there are archives. Writing and publishing is an old technology, and the older a technology is, the more likely it is to survive.

Ghosts in the Internet Archive: As you might expect, another place where ghosts of a digital project live on is in the Internet Archive. I didn’t count these archived copies of an online project to mean that they were “accessible” online, by the way, but for projects that did provide a URL, this was a place to find more information about the project’s life and afterlife.
Libraries, preservation, & weeding

When I titled this talk many months ago, I hadn’t yet investigated the role of the library in these DH 2005 projects. Turns out, it’s very small. There was not much library involvement in these projects, beyond research help or usage permissions. Some presenters were librarians. Only two projects, Mayors and Sheriffs of London and the Carte Calendar Project, were hosted on an academic library’s web space (and, I’ll note, they continue to be). There are many more DH centers and projects that have partnered with libraries since 2005, so the library factor will be something to be studied in the future. After all, libraries are very adept at acquiring and preserving materials. It is what we do.

Is it so bad to let a digital project wither and die? Some might have been experiments or pilot projects, not worth keeping around, in terms of significance or viewership. In libraries, the practice of weeding a collection is essential: if a book is outdated or unused, it is discarded. The disappearance of these online projects may be an organic form of weeding, with or without a human agent. But looking at the DH2005 data, some of the projects that are no longer available were expensive, labor-intensive efforts. One project was supposed to be one result of a $100,000 NEH grant and a $50,000 Mellon grant. I’m not going to point out any in particular, but it seems quite a shame to lose such large scholarly efforts.

Conclusions

Having looked at the survival statistics of digital humanities projects from ten years ago, and seeing that only half are still online, we might understand a little more about the nature of digital work.

In this panel, we’ve seen presentations that integrate the very old and the very new. Printed texts from the 17th century; catalog records from the early 20th century; code from last year. This is what is so engaging about digital work: you’re looking through time, using tools built in different centuries, to follow a question.

Walt Disney Studios.
Think of it perhaps using the metaphor of the multiplane camera, where the background of the image is at the very bottom, with layers of foreground stacked up in between. With layers of images printed on transparent cels, you set a scene that you can view all the way to the bottom, to the source.

Fleischer studios. Found on How Stuff Works.

The purpose of the multiplane camera is to produce correct perspective in animated sequences. The background is almost static, and the foreground flashes by.

Walt Disney studios. Found on Giphy.

Or, zooming in, the foreground falls away as the background comes into focus. Or, zooming out, the foreground becomes part of the background.
What stays in the frame? What remains visible, accessible? An early modern printed text is hundreds of years old, and given a safe place to live, it will remain with us for hundreds more years. The ESTC is decades old and stable. A digital resource from DH2005, say, might or might not be around, there’s a 50/50 chance. A digital work from yesterday — who knows? It may sink into the background of stable survival. Or it may flash through the frame, as the detritus of digital scholarship.

In digital work, we use layers of media and remediations; we repurpose things built by others; projects run through many versions, building layers of themselves; and we must also address layers of time. Digital work is multiplanar time-based media. Untouched, the life of a digital project will be animated over time with transformations and disappearances of its many layers. When the sources you use and reuse and cite die slowly or quickly, what does that mean for your claims, for the reproducibility of your results? Preservation work begins at the moment of creation. If your own digital project has a short projected half-life out the gate, would that change how you produce and share it? As libraries, archives, and other repositories of our scholarly and cultural heritage address digital preservation problems, scholars who do digital work must be invested in preserving their work, too — or we run the risk of losing an entire strata of scholarship.

Robin Camille Davis
2015

Posted online at robincamille.com/presentations/mla2015/