**Endings: Concluding, Archiving, and Preserving Digital Projects for Long-Term Usability**

**By Emily Comeau**

**Introduction**

This paper is an overview of a study conducted by the Humanities Computing and Media Centre at the University of Victoria as part of the SSHRC-funded (Social Sciences and Humanities Research Council of Canada) project *Endings: Concluding, Archiving, and Preserving Digital Projects for Long-Term Usability*. I will start with a brief explanation of the goals of the *Endings* project, which will be followed by an overview of the four flagship *Endings* case studies, as well as the subsequent qualitative study. I will then discuss some of the initial results from this research, both the positive and negative experiences that DH scholars around the world have shared with us. Then I will share a number of recommendations from the *Endings* team, which are the culmination of this study itself as well as our own experiences through the four case studies.

**What is the *Endings P*roject?**

The *Endings* project developed from a growing awareness in the DH community that digital humanities (DH) projects were not being preserved after their end-dates. Rather, they were degrading and disappearing. This led us to ask: how do DH projects end and where should they be archived in order to maintain their dynamic, interactive features? Beyond this, we ask: what factors ensure the survival or precipitate the disappearance of DH projects?

To explore these questions, we have developed and reflected on four case studies, four projects conducted by *Endings* team members (GitHub page <https://github.com/projectEndings/Endings/wiki>): *Le Mariage sous L’Ancien Régime*, the *Map of Early Modern London*, *the Nxa'amxcín Database/ Dictionary*, and *Robert Graves’ Diary*. These four projects have led to the development of techniques, tools, and approaches for ending and archiving which can be applied to other DH projects.

**Le Mariage sous L’Ancien Régime** (<https://mariage.uvic.ca/>), with Dr. Claire Carlin

*Le Mariage sous L’Ancien Régime* is a digital anthology of early modern (mid 1500s to early 1700s) French texts surrounding the topic of marriage. It includes:

* Annotated satirical texts, plays, novels, treatises, and other documents;
* Links to an extensive index from historical figures, characters, and places within texts;
* Short commentary in the form of numbered notes on each text, which “provide commentary on ambiguous, little known or difficult concepts;”
* Longer commentary in the form of articles which link to the texts and images;
* “An extensive bibliography;”
* Images which have been annotated and made searchable using the Image Markup Tool, as developed by Martin Holmes.

Work on *Mariage* is ongoing, as we continue to proofread and markup texts, add annotations, and improve site functionality. (ref: the GitHub page)

**Map of Early Modern London** (<https://mapoflondon.uvic.ca/>), with Dr. Janelle Jenstad

The Map of Early Modern London, or MoEML, is really “four distinct, interoperable projects: a digital edition of the 1561 Agas woodcut map of London; an Encyclopedia and Descriptive Gazetteer of London people, places, topics, and terms; a Library of marked-up texts rich in London toponyms; and a versioned edition of John Stow’s Survey of London [from 1598]” (ref: MoEML project page). MoEML is ongoing, and recently received an influx of SSHRC funding, which will support the next phase of project development.

**Nxa'amxcín Database/Dictionary**, with Dr. Ewa Czaykowska-Higgins

The Nxa'amxcín Database and Dictionary project is a unique project, in that it “is [the] only lexical resource in the world that uses TEI,” and because it brings up usability issues between electronic vs. print dictionaries that “haven’t been explored [much] by linguists.” It is also “the only written record of the language” (ref: Endings GitHub page). The project itself is only accessible to members of the Nxa'amxcín community, as per the community’s wishes. It is for this reason, as Dr. Czaykowska says, that “conversations about preservation and access will be challenging” (ref: GitHub). There are, however, a number of reports and articles that have been published about this project (“Using TEI…” paper: <http://nflrc.hawaii.edu/ldc/ldc-vol-8-2014/>). Work on migrating, encoding, and updating linguistic analyses is ongoing, as well as discussions about long-term maintenance and preservation of the project in general.

**Robert Graves’ Diary** (<http://spcoll.library.uvic.ca/Digit/GravesDiaryProject/home.htm>), with Dr. Elizabeth Grove-White

The *Robert Graves’ Diary* project is an archive of diary entries written by Robert Graves, including letters, photographs, newspaper clippings, etc. “The project’s objective is to produce the first scholarly edition, in print and electronic form of this unpublished diary.” The goals of this project are:

* To provide internet access using XML technologies to this unique text by Robert Graves;
* To encode the diary text electronically with supporting stylistic, interpretive and contextual markup useful for researchers;
* To allow for complex (although currently limited) searching of the diary text via the internet using an XML database eXist;
* To give researchers a powerful text analysis tool, and to establish a prototype for future electronic publishing and archiving projects. (ref: the RGD project page)

Active development of this project ended in 2004, but work on maintenance and preservation continues at the UVic HCMC.

Each of these projects has presented its own challenges for preservation and maintenance of functionality, and has led to the development of markup and archiving tools, data modelling techniques, and preservation standards for DH projects. In addition to reflecting on our own projects, we have also conducted a qualitative study, which comprised a survey and interviews with DH scholars around the world. We are currently in the process of analyzing both the survey and interview data, and this paper will discuss some preliminary findings from our analysis of the interviews.

**Methodology**

The following discussion of our research methodology demonstrates our overall approach to this research, particularly the interview portion of our study, including the data collection, organization, and analysis. Methodology in this context is not only the philosophical foundation for this portion of the *Endings* project (qualitative research), as well as the framework for developing specific methods to be used for data collection and analysis (interviews, coding, etc.), but it is also the basis for establishing how we as researchers are situated within this research. This study follows a constructivist grounded theory approach, using elements of both narrative inquiry and phenomenology to understand and interpret the experiences of interview participants.

In a constructivist grounded theory (CGT) approach, the role of the researcher is to co-construct meaning with participants (through interviews), and then to reconstruct and interpret meaning from the data, "striving to know the world from the viewpoint of the research participants" (Mills et al. 2014, p. 7). According to Mills et al. (2014), "it is impossible to separate researcher from participant in the generation of data" (p. 7). It is in this way that researchers are also participants in the study. A key aspect of a CGT approach is researcher reflexivity. As Mills et al. (2014) explain, "researchers must think about what they themselves are doing, be explicit about how and why they are doing it and consider the effect they are having on the data and eventual findings" (p. 7). In a CGT approach, data is not removed from participants, but rather interpreted and explained in the context of participants' voices. CGT is generally intended to inductively generate theory, as "an explanatory scheme comprising a set of concepts related to each other through logical patterns of connectivity" (Mills et al. 2014, p. 8). However, since we are not intending to generate a "theory," we seek rather to "[employ] a ‘grounded theory approach’ ” (Mills et al. 2014, p. 8).

Narrative inquiry and phenomenology both overlap with CGT in a number of ways. In narrative inquiry, the researcher is expected to immerse themselves in the topic or world that they are studying. According to Lewis (2014), "narrative work entails intimate engagement of the researcher and the co-researcher(s) in the story of the relationship. You write yourself into the work" (pp. 5-6). In a narrative inquiry approach, "authentic understanding is gained from the meanings people bring to phenomena” (Lewis 2014, p. 4), and these phenomena are best studied "in their natural settings" (p. 5). Narrative inquiry focuses on uncovering meaning "rather than measurements" (Lewis 2014, p. 10), and often employs specific methods such as conversations and interviews that include the researcher as participant. Rather than relying on statistics or numerical results, a qualitative analysis that uses a narrative approach is based on "evidence, interpretive plausibility, logical constructions and disciplined thought" (Lewis 2014, p. 10).

According to Usher & Jackson (2014), "phenomenology is the careful and systematic reflective study of the lived experience" (p. 2), which "aims to bring deeper insight" to real-world phenomena. A phenomenological approach necessitates reflexivity on the part of the researcher, "in order to understand the nuances of your own experience/s of a phenomenon and then to appreciate the subtleties of being a person experiencing the phenomenon in a specific context” (Usher & Jackson 2014, p. 8). In a phenomenological approach, researchers "become the instrument of the research" (Usher & Jackson 2014, p. 8) and are responsible for interpreting the experiences shared by participants as co-researchers. It is in this way that knowledge is co-constructed by researcher and participant.

Initial data collection was conducted through a survey that we built on the LimeSurvey platform. At the end of the survey, respondents were given the option of participating in a semi-structured interview. The formal analysis of this aspect of our research (survey and interviews) is currently underway, with results just beginning to emerge. However, for the purposes of this paper, I have conducted a preliminary analysis of the transcribed interviews. In the next section, I will discuss the major themes that have emerged in initial close readings of these transcriptions.

**Results**

A number of the scholars we interviewed shared stories with us detailing their successful project endings. In my initial analysis of the interviews, several common themes have emerged in terms of what contributed to the success of these project endings. These themes are as follows:

1. Support in the form of funding, expertise, technological support and hosting, and infrastructure from individual institutions or national public organizations. This also includes support in the form of standards of practice and infrastructure within communities of practice.
2. Separating content, structure, and processing from each other, or separating data and platform, and preserving each separately.
3. Pan-institutional partnerships and multi-institutional collaboration, at both the national and international level.
4. Consistency in coding and markup procedures and fastidious recording of both data and metadata.
5. Agreements between parties, including between scholars, but especially with hosting institutions, that clarify and guarantee responsibility and terms of support.
6. Storing multiple versions in different places.
7. Using software that is stable, supported, and easily maintained or updated to current versions.
8. Preserving what needs to be preserved in a way that supports the goals of the project. In some cases, it was neither prudent nor necessary to create an archive of the project. Rather, in terms of knowledge mobilization, it was more important that the outcomes of the project be shared through reports and presentations. In some cases, other projects have evolved from the original project.

Many scholars have also shared with us their stories of unsuccessful endings, or their “disaster stories.” In many cases, our interviews seemed to function as sort of “therapy sessions,” where scholars felt comfortable sharing their frustrations about endings gone wrong. The following themes bring to light factors that have contributed to the failure of these project endings:

1. A lack of resources for archiving and preservation, including funding and expertise.
2. A lack of institutional support and integration. For instance, sometimes the project is not valued at an institution because DH is not valued. In some cases, there has not been adequate infrastructure at an institution to support the preservation of DH projects. In other cases, the institution was supportive when the project started, but a change in institutional leadership led the institution to withdraw their support.
3. The lead scholar leaves the institution or dies and the project is no longer accessible to other team members. In some cases, the institution has locked down access to the project for the other team members, particularly if they are at different institutions. In other cases, all team members left the institution, and eventually the hosting server broke or was taken offline, or the institution simply decided to stop hosting the project, and there was no-one left at the institution to advocate for the project.
4. There is no licensing agreement, and so ownership of the project, as well as hosting and maintenance responsibilities are unclear. In some cases, either the institution or the primary investigating scholar is proprietary and will not share ownership or access with other scholars or institutions.
5. There has been no effort to develop a preservation or sustainability plan.
6. Hosting is fragile, particularly when there is no agreement or understanding of responsibilities between scholars and institutions.
7. The project is complicated and involves multiple complex parts that each need to be preserved in different ways. This usually requires a level of technical expertise or access to software or techniques that are not readily available to the project team.
8. The project has been built on software or applications that are not being supported long-term and that cannot be maintained without consistent staff and resources.

Overall, the biggest issue we have heard throughout the interviews is simply not planning for the end of a project. This often involves asking questions of our co-investigators as early as possible in the project that we don’t like asking, such as “what happens to the project if one (or both/all) of us leaves or dies? Will we still want to maintain this project five, ten, or even twenty years after active development ends? What do we do if the worst happens, such as if the server breaks, or if the software is no longer supported, or if funding runs out, etc.?”

**Our solutions**

Through our experiences with ending and archiving numerous DH projects, we have compiled a set of suggested milestones for making a project “archive ready,” which should be addressed during the planning stages of a digital project. Our interviews with other DH scholars have not only reinforced these suggestions, but have also expanded and enriched our recommendations for ending and archiving DH projects. As *Endings* team member Martin Holmes explains, a project that is ready to be archived is:

* functional without a server infrastructure to support it;
* dependent only on existing standards with massive support (HTML5, JavaScript, CSS, XML);
* harvestable by archiving tools;
* functional when harvested and served up by an archive browser such as the Wayback Machine;
* search-engine-friendly;
* metadata-rich, preferably in multiple formats.

(ref: “What does it mean for a project to be "archive ready"?” <https://github.com/projectEndings/Endings/wiki/What-does-it-mean-for-a-project-to-be-%22archive-ready%22%3F>)

Project plans should take into account the resources necessary for maintaining both data hosting and usability long-term. For instance, what applications or software can be implemented now that will still be usable (in their current iterations) in the future? Moreover, it is important to consider what aspects of a project should be and can be preserved or archived, and the most realistic methods for doing so. Related to these considerations, it is also important to keep in mind the original purpose of the project. Digital editions, for example, will have different preservation needs from gazetteers, user platforms, or outcome-based projects, and so on. As such, not every project will benefit from the same approach to ending and archiving.

At this point in our research, having conducted a preliminary analysis of the interviews, and reflecting on our own experiences of ending and archiving digital projects, we have compiled a number of concrete strategies for long-term preservation:

1. Use well-supported and easily flattened markup languages such as XML and HTML. Following the standards of organizations such as the TEI can also provide standards and infrastructure that are invaluable for preserving DH projects.
2. Preserve copies of the project in multiple repositories. It can also be beneficial to preserve parts of the project separately, e.g. store the code in GitHub (or a similar repository) and project outputs in other accessible databases.
3. Make sure any websites are crawled and dumped regularly. Also, check periodically that the site is actually functional on the Wayback Machine or Internet Archive. Mirror sites are a useful strategy for ensuring that a project website exists in multiple locations online.
4. Develop a strategy early on for storing and preserving non-text media such as images, audio, mapping, etc. For instance, the HCMC has developed an Image Markup Tool (<https://tapor.uvic.ca/~mholmes/image_markup/>) which has made it possible to create markup for images in projects such as *Le Mariage sous L’Ancien Régime.*
5. Consider what aspects of the project you hope to preserve, or that must be preserved. For instance, do you hope to preserve the project website, or is it more important to preserve the data and outcomes of the project regardless of the website? Or is the purpose of the project a digital archive that includes an attractive user interface? Asking these kinds of questions can help to establish goals for long-term preservation and help focus a project. (ref. to Lynne Siemens’ work on project management?)
6. Establish formalized agreements with hosting institutions that clearly state data ownership, kinds of support and resources provided by all parties, and responsibilities for project maintenance in short-, middle-, and long-term contexts.

**Conclusion**

In many of the experiences of unsuccessful endings that fellow scholars have shared with us, it has been evident that those involved have done everything right, and that something unpredictable or unprecedented has happened, which caused the project to be lost. There is no way of knowing that these events will take place. What we hope to be able to offer the DH community through the *Endings* project is a set of resources, a toolkit, for how to plan for the end of a project, and to be prepared for any situation that may arise so that if the unthinkable happens, the project is still safe and preserved, or at least retrievable.

Future publications from the *Endings* team will discuss our full analysis of both the interviews and the survey responses, and will further elaborate on our recommendations for ending and archiving DH projects. Finally, alongside the results of this study, the toolkit that we develop will also “address intellectual property concerns along with maintenance strategies, long-term stability, and access issues. We will consider the needs of various stakeholders and potential repositories, including universities, memory institutions, non-profits, and national and international governments” (<https://github.com/projectEndings/Endings>).

**Acknowledgements???**

**About the Endings team:**

This research is being conducted by a multidisciplinary team of ten investigators at the University of Victoria: Stewart Arneil, Martin Holmes, and Greg Newton are developers in the UVic Humanities Computing and Media Centre; Matt Huculak, John Durno, and Lisa Goddard hold positions in the UVic Library; Claire Carlin, Ewa Czaykowska-Higgins, Elizabeth Grove-White, and Janelle Jenstad are faculty members. Our team has collective theoretical and practical experience in geohumanities, multimedia markup, endangered language documentation, textual editing, digital publishing, and text encoding. This paper has been presented by *Endings* research assistant Emily Comeau, who has been involved in the project since 2017, most recently in co-ordinating, transcribing, and analyzing the interviews. Emily has recently completed an MA in Linguistics at the University of Victoria, where her research explored the role of literacy in Indigenous language revitalization and the implications of print literacy in the context of decolonizing education.