

---

## Shelf life: Identifying the Abandonment of Online Digital Humanities Projects

Luis Meneses

ldmm@uvic.ca

Texas A&M University, United States of America

Richard Furuta

furuta@cse.tamu.edu

Texas A&M University, United States of America

---

The Internet, along with the advent of online technologies has provided researchers with greater opportunities to collaborate and create a myriad of digital projects. Taking our research group as an example, we have collaborated in creating online art catalogs (Meneses et al., 2011), interfaces for visualizing and creating poetry (Meneses et al., 2013), and tools for analyzing and exploring Shakespeare's plays (Meneses et al., 2015, Meneses et al., 2016a). However, the convenience and familiarity of computer networks makes us forget (or overlook) that there is a certain fragility associated with our online tools. In turn, this fragility threatens the completeness and the sustainability of our work over time.

Nowadays, a large portion of the research carried out in the digital humanities includes an online project as one of its components. In turn, these digital projects can be catalogued as distributed resources, which implies that the administrative control of information related to a topic may be spread across online resources and/or collections maintained by multiple scholars in different institutions. This administrative decentralization can lead to changes in content that are often unexpected by a researcher.

These unexpected changes can be caused by different factors or circumstances. Changes can occur because of deliberate actions on part of the collector – for example, reorganization of the structure of the collection, switching to a different content management system, or changing jobs and institutions. Changes might also be due to unexpected events – earthquakes, power outages, disk failures, – or may be due to other uncontrollable factors – death, seizure of computers by law enforcement, or

termination of the services from an Internet Service Provider (Mccown et al., 2009).

Over time, great strides have been made to harness and manage the fragility of online resources. Klein and Nelson argued that digital documents do not disappear from the Web, but leave artifacts that can be used to reconstruct them (Klein et al., 2011). Bar-Yosseff et al. carried out experiments to measure the decay of the Web (Bar-Yosseff et al., 2004). SalahEldeen determined that nearly 11% of shared resources will be lost one year after being published and that this decay will continue at a 0.02% rate per day (Salaheldeen and Nelson, 2012). Nevertheless, and despite these previous efforts, managing and characterizing change in online environments is a complex problem.

Recently, our research group has been focusing on analyzing the perceptions of change in distributed collections (Meneses et al., 2016b). However, we believe that the inherent characteristics of online digital humanities projects present an interesting (and unique) area for inquiry for two reasons. First, the research aspect of digital humanities projects hinders our previous approaches – as our methods for identifying change in the Web do not fully apply. And second, digital humanities projects have a limited useful life – which is accompanied by research from primary investigator, which may or may not be indicated by updates in the project's content and tools. We have seen many cases of successful projects in digital humanities (that fulfill their original objectives and achieve their expected level functionality) that interestingly become abandoned at some point in time. Examples of abandoned successful projects include the [Cervantes Project](#) and the [TAMU Herbaria Project](#). This abandonment might be caused by a different set of reasons –which are not often apparent to its users– such as loss of funding, change in personnel or simply decay in interest. We believe that all these reasons are worthy of study.

All this reasoning led us to formulate the following question: When can online digital humanities projects be considered abandoned? In this paper, we propose to present a study on the persistence and average lifespan of online projects in the digital humanities. More specifically, we will elaborate on their reliance on distributed resources and methods for measuring their shelf life: the average length of time that a digital project can endure without updates until it can ultimately be considered abandoned by its researcher.

Furthermore, we believe that “abandonment” is not necessarily a sufficient designation —as there are

different nuances involved. We will proceed to elaborate on them using one of our online projects as an example. [Digital Acting Parts](#) is an online project that encourages active reading and memorization, which in turn leads to a better understanding of Shakespeare's plays. The project has been active since 2013, but online development has shifted to a set of different processes that are carried out behind the scenes. Consequently, the project's online presence has not been updated for some time now (we estimate that it has been at least a year). However, the online tools are quite stable at this point. In this specific case, the lack of updates and new content is not a signal of abandonment. This is clear example of why the rules for traditional websites do not fully apply and new metrics are needed to identify issues concerning online projects in the digital humanities.

Our study is an attempt to categorize change in a very specific domain. As an attempt of categorization, determining the degree of abandonment affecting a digital project over time is a difficult task. A Web resource may gradually degrade from being correct to one that is still of some use by providing access to related information or information about the institution to contact for more information. Abandonment can also be hinted by changes in Web servers, directory structures, etc., which may cause Web requests to still result in a successful responses from a server, yet provide no valid information to the requestor. Based on our findings, we approximate the average shelf life to 5 years, which aligns with reports from previous work (Goh and Ng, 2007).

Additionally, our study will touch upon on potential strategies for the archival and the long-term preservation of abandoned digital online projects. It is important to highlight that different levels of preservation and curation are needed among digital projects. Historically, preservation efforts have been primarily concerned with maintaining the primary artifacts in collections; relegating descriptive metadata to a lesser level of importance. There is an underlying notion that descriptive metadata is static: requiring minimal resources to maintain and consequently making it easier to preserve. However, our previous work (Meneses et al., 2016b) has shown us that this is not the always the case.

To summarize, in this paper we propose to identify indicators of the abandonment of digital humanities projects – as well as identifying their average lifespan. Digital online projects in the humanities have unique characteristics that make them impervious to the metrics that used in the Web as a whole. In our

opinion, these unique characteristics make them worthy of study. In the end, the purpose of our study is to gain a better understanding of digital humanities projects, their lifespan and formulate better strategies for their long-term preservation.

## Bibliography

- Bar-Yossef, Z., Broder, A. Z., Kumar, R. & Tomkins, A.** (2004). Sic transit gloria telae: towards an understanding of the web's decay. *Proceedings of the 13th international conference on World Wide Web*, 2004 New York, NY, USA. ACM.
- Goh, D. H. L. & Ng, P. K.** (2007). Link decay in leading information science journals. *Journal of the American Society for Information Science and Technology*, 58, 15-24.
- Klein, M., Ware, J. & Nelson, M. L.** (2011). Rediscovering missing web pages using link neighborhood lexical signatures. *Proceedings of the 11th annual international ACM/IEEE Joint Conference on Digital libraries*, 2011 Ottawa, Ontario, Canada. ACM.
- Mccown, F., Marshall, C. C. & Nelson, M. L.** (2009). Why web sites are lost (and how they're sometimes found). *Communications of the ACM*, 52, 141-145.
- Meneses, L., Estill, L. & Furuta, R.** (2015). Digital Acting Parts: Learning and Understanding Shakespeare's Plays. *Joint CSDH/SCHN & ACH Digital Humanities Conference 2015*. Ottawa, Canada.
- Meneses, L., Estill, L. & Furuta, R.** (2016a). This was my speech, and I will speak it again": Topic Modeling in Shakespeare's Plays. *Joint CSDH/SCHN & ACH Digital Humanities Conference 2016*. Calgary, Canada.
- Meneses, L., Furuta, R. & Mandell, L.** (2013). Ambiances: A Framework to Write and Visualize Poetry. *Digital Humanities 2013*. University of Nebraska-Lincoln.
- Meneses, L., Jayarathna, S., Furuta, R. & Shipman, F.** (2016b). Analyzing the Perceptions of Change in a Distributed Collection of Web Documents. *Proceedings of the 27th ACM Conference on Hypertext and Social Media*, Halifax, Nova Scotia, Canada. ACM, 273-278.
- Meneses, L., Monroy, C., Furuta, R. & Mallen, E.** (2011). Computational Approaches to a Catalogue Raisonné of Pablo Picasso's Works. *Interdisciplinary Journal for Germanic Linguistics and Semiotic Analysis*.
- Salaheldeen, H. M. & Nelson, M. L.** (2012). Losing My Revolution: How Many Resources Shared on Social

Media Have Been Lost? *Proceedings of Theory and Practice of Digital Libraries 2012*, 2012 Paphos, Cyprus.