The Archive as Repertoire: Transience and Sustainability in Digital Archives

Miguel Escobar Varela <contact_at_miguelescobar_dot_com>, National University of Singapore

Abstract

Digital archives change more quickly than traditional ones: they are adaptable and transient. This has advantages and disadvantages; digital archives can disappear from sight almost instantly but they can also be easily safeguarded and restored. Borrowing the critical vocabulary of performance studies, digital archives could thus be understood as "repertoires" rather than traditional archives. By treating digital archives as repertoires, this article explores different threats and opportunities presented by their volatile nature and makes policy and technical recommendations on how to ensure their relevance and sustainability.

Introduction

"At the moment this archive is only a prototype, but we will soon have a public version available online." I have heard – and pronounced – these words in several occasions in digital humanities conferences and presentations. Everyone seems to be building an archive, but these projects are often archives in progress. And the most established archives also change often: new versions, features and editions are constantly added. When you visit the website of many archives, it is common to find features that are “not yet” working or to read in the history of the archives about aspects of the archive that were present in previous iterations of the projects.

It is easy to condemn this situation. Traditional archivists could assert that digital archives are not really built to endure the passage of time and constitute a repository for future historical research. Therefore, the argument goes, digital archives are not really archives. However, comparing digital with non-digital archives obviates many of the specific ways in which digital archives can be relevant and sustainable, two important objectives of every archiving project. There are two complicating factors to an easy comparison with non-digital archives:

- The mutability of digital archives is complex.
- The lifespan of a digital archive is indeterminate. Any iteration of the archive can disappear quickly, but it can also be reconstructed in ways that are difficult – if not impossible – with brick and mortar archives.

This article proposes to address these complications through the theoretical lens of performance studies, a discipline that has long engaged in the theorization of archives. In performance studies, the opposite of an archive is a repertoire. And though there are fundamental differences between digital archives and embodied performance repertoires, there are also striking similarities between them. Looking at the way performance scholars conceptualize and study repertoires, digital humanists can gain a strategic new vocabulary to describe how digital archives work and, most importantly, think about ways to upkeep and maintain the relevance of archives in the future. Before launching into a performance studies approach, this article will briefly consider how the differences between digital and traditional archives have been conceptualized earlier. This is a long and complex discussion, but several key issues are constantly highlighted in the conceptualization of digital archives: the possibility of new methodological approaches, the threat of the disappearance of context, the new need for new technical models, the competing claims of access and preservation, and the suitability of the word “archive.”
Much has been said about the capacity for search-and-retrieve and data mining operations allowed by digital archives. One of the key areas where these techniques are already making great an impact is newspaper archives and the historical research they enable. Newspaper archives are unique because of the periodicity, abundance, and uniformity of their records; thus, they provide us with an excellent case study for the impact of new methodologies for archival research. According to Bob Nicholson, who writes about the impact of digital archives in the study of newspapers, everything has been changed by search-and-retrieve as well as data-mining algorithms which could not have been applied to physical newspaper collections. New technologies are affecting a broad range of questions, whether it is to study the lives of individuals, explore changes in the usage of language, track the development and circulation of ideas, or write the history of a serialized publication:

Digital methodologies promise to have a significant impact on multiple areas of historical research. We are potentially on the cusp of a [...] revolution, a ‘digital turn’ in humanities scholarship driven by the creative use of online archives and a willingness to imagine new kinds of research. [Nicholson 2013, 63]

Archives dealing with literature, theatre and other arts are usually not as uniform or as extensive as newspaper archives. Nevertheless, different ways of interrogating archives are also changing the ways all kinds of digital archives are conceptualized and used. Another concern often brought up in discussions about digital archives is the disappearance of context. Discussing 19th century newspaper and literature archives, Shafquat Towheed remarks that the context of reading was different in the 19th century and the specificity of reading is difficult to reconstruct from digital collections, since they generally lack annotations and marginalia. Another problems is that items are not bundled together, which makes the practice of browsing through the materials cumbersome:

Paradoxically, while the digitization of nineteenth-century newspapers means that they have never been more widely available, the very experience of casually browsing through the pages of an essentially disposable publication has become increasingly remote and difficult to reconstruct. [Towheed 2010, 142]

Computer scientists are also trying to come to grips with the different conceptual model required by digital archives, as opposed to that of digital libraries or other kinds of collections. Following Stefano Vitali, Nicola Ferro and Ginamaria Silvelo suggest that the most distinctive characteristic of digital archives is that they are held together by an archival bond: there is a specific, homogeneous and often serialized way in which items in a collection are organized and related to one another. Ferro and Silvelo propose a formal model to address these issues, NEsted SeTs for Object hieRarchies (NESTOR) [Ferro and Silvelo, 2013]. However, the development of information organization models is still an open question and there are currently no standards to be widely adopted.

Another challenge of digital archives refers to the competing claims of access and preservation. Should an archive emphasize the widest range of materials possible? Or should it make sure that a smaller collection is properly documented, annotated, and made accessible to non-specialists? Different archives have different philosophies on this matter, but, as Michelle Warren notes, every cultural record, including books, is shaped by this competing claims:

The effort to preserve and the desire for access have both changed the text now recorded. Together, preservation and access make archives malleable and dynamic rather than static. [Warren 2014, 170]

Another key concern is the suitability of the word archive. Kenneth M. Price looks at different words that are used in digital humanities work: edition, project, database, archive, and thematic research collection, before suggesting a new potential term: arsenal. When discussing all of these terms, he stresses the importance of the choice of words and also notes the advantages and limitations of all the possible choices. His characterization of digital archives is particularly insightful:

In a digital environment, archive has gradually come to mean a purposeful collection of surrogates. As we know, meanings change over time, and archive in a digital context has come to suggest
something that blends features of editing and archiving. To meld features of both — to have the care of treatment and annotation of an edition and the inclusiveness of an archive — is one of the tendencies of recent work in electronic editing. [Price 2009]

Whether or not another word becomes more common in the future, archive is the currently the most common term to describe the digital artifacts I have in mind. This is perhaps how most projects address the competing claims of access and preservation.

The previous overview has shown some of the main concerns and anxieties relating to digital archiving: access and preservation, methodological implications for research, formal computer-science models, the disappearance of context, and the suitability of the word archive.

In the rest of this article I will use terms prevalent in performance studies – and suggest new concepts – that can help us understand the constraints and possibilities of digital archiving. All of the issues mentioned above can be traced to the dynamic, changing nature of archives. Although this dynamism is often mentioned, I argue that the particular way in which digital archives are dynamic demands further attention. Based on the terminology of theatre studies and my own experiences working with digital archives, I propose two new terms: internal dynamism and external dynamism in order to qualify the ways that digital archives change.

This article is organized as follows: the first section offers a definition of archives and repertoires from performance studies and explains why digital archives can be studied and managed as repertoires. The second and third part look at the internal and external adaptability of archives. The last two sections offer recommendations on how digital archives can be better managed to ensure sustainability and relevance. For this, the article borrows tropes and techniques from the open-source software development culture.

**The conceptual tools of performance studies**

Having described the complex nature of digital archives, I suggest that performance studies can offer a new way of understanding the connection between archives, memory and knowledge. Performance studies conceptualize archives in two major ways: as a metaphor and as a problematic source of historical knowledge about performance practices. Metaphorically, the archive represents the way in which certain theater makers work. For example, Mike Frangos argues that Samuel Beckett's plays are archival in the sense that they enact, in form as well as in content, the main concerns of archiving: historicity, repetition, memory, and erasure [Frangos 2012, 217]. This claim is similar to that of Kenneth M. Price, who argues that the work of Walt Whitman was developed, metaphorically, as a database. Whitman constantly republished Leaves of Grass, rearranging verses drawn from a vast personal repository: “his storehouse of poetic lines, in both manuscript and print, was his working database for future compositions” [Price 2009].

The archive can also be a metaphor that stands in for any theatre performance. In this sense, theatre is always a living, evolving archive where things are changed as they are stored and retrieved. Marvin Carlson describes theatre as a place where ghosting occurs. By this he means that everything in the theatre has been “used before”. The same bodies of the actors have interpreted other roles, the same words have been spoken and the same objects have been used to convey other meanings [Carlson 2003]. Following the example of theatre studies, every aspect of cultural production – anything and everything in a digital humanities archive – could be understood in terms of archival metaphors and ghosting. As Jones et al. assert:

> Arguably, all archiving is performance: records are surrogates that provide a window onto past moments that can never be recreated, and users interact with these records in a performance to reinterpret this past. [Jones et al. 2009, 166]

However, a more interesting avenue comes from the second conceptualization of archiving in performance studies, where archives – in a more conventional sense of the word – are seen as the problematic repositories of theatrical items: photographs, booklets, recordings, technical scripts, and interviews. Whether these items are ephemera or documentation records, they are problematic in the sense that they are not performances. A film archive will contain the
films, and a literary archive will contain manuscripts. But a theatre archive will never contain theatre performances. Theorists of theatrical archiving have thus always dealt with a pressing concern of digital archives: the fact that the items in a collection are not the “real” things but mere surrogates, data, and traces. Peggy Phelan has famously argued that “performance becomes itself through disappearance” and that the records of a performance can never be equivalent to the performance itself:

Performance's only life is in the present. Performance cannot be saved, recorded, documented, or otherwise participate in the circulation of representations of representations: once it does so, it becomes something other than performance. To the degree that performance attempts to enter the economy of reproduction it betrays and lessens the promise of its own ontology. [Phelan 1993, 146]

This argument has been contested by Phillip Auslander, who argues that performance and mediatized recordings cannot be distinguished on an ontological basis:

Mediatized forms like film and video can be shown to have the same ontological characteristic as live performance, and live performance can be used in ways indistinguishable from the uses generally associated with mediatized forms. Therefore, ontological analysis does not provide a basis for privileging live performance as an oppositional discourse. [Auslander 1999, 184]

In her 2010 book on Cyborg Performance, Jennifer Parker-Starbuck historicizes this controversy and labels it a “dated” disagreement, suggesting that history has sided with Auslander and confessing to a “tacit acceptance of Auslander's argument” that the live is already mediatized “in the contemporary moment of globalized technology” [Parker-Starbuck 2011, 9]. Nevertheless, the argument brought forth by Phelan still stands and should still be relevant to digital archivists. Even if the “mediatized” is now irremediably part of the “live”, there is still an ontological difference between a live performance and a record of that performance. This difference might not grant a political superiority to one of them (as Phelan expected), but it should still inform the efforts of digital archivists. It is important for us to recognize the differences between the digital surrogates in an archive and the original sources. This necessarily calls for specific practices of contextualization, such as explaining how the records were made and what gaps are present in a documentation project.

The theatrical archive is also problematic because it stands in opposition to another mode of transmission through time and space, the repertoire. The repertoire is an archive of the body and of oral knowledge. The differences between both modes of transmission have been thoroughly explored by Diana Taylor in the landmark The Archive and the Repertoire. Consider, for example, the way in which she describes archival memory:

Archival memory works across distance, over time and space [...] what changes over time is the value, meaning or relevance of the archive. Insofar as it constitutes materials that seem to endure, the archive exceeds the live. [Taylor 2003, 19]

This kind of memory is set in opposition to the one enabled by the repertoire:

The repertoire, on the other hand, enacts embodied memory: performances, gestures, orality, movement, dance, singing - in short, all those acts often thought of as ephemeral, non-reproducible knowledge. The repertoire requires presence; people participate in the production and reproduction of knowledge by “being there”, being a part of the transmission [...] The repertoire both keeps and transforms choreographies of meaning. [Taylor 2003, 20]

In terms of what they hold – materials that are meant to endure distance in time and space – digital archives are no different from conventional ones. However, as Jones et al. notice, code-based archiving does require interaction and the presence of the users:

Digital records are inherently performative, only coming into existence when the correct code executes the data to render a meaningful output. [Jones et al. 2009, 170]
However, despite the performative aspect of the computer code they are built upon, digital archives don't embody ephemeral knowledge that requires co-presence. They require interaction but not the embodied co-presence of producers and consumers of meaning. Digital archives do behave as repertoires in that they are constantly evolving and they also keep and transform “choreographies of meaning”, in Taylor's evocative phrasing. This article suggests that there are two ways in which this similarity with repertoires is manifested: constant change that is driven by improvisation and the possibility of embedding the items of the repertoire in entirely new settings.

In what follows I suggest that the first of these characteristics is equivalent to the “continuous beta” status of many digital archives; and the second, to the potential to embed archival content in new settings through iframes and APIs.[1] These characteristics correspond to two kinds of dynamism, which I propose to call internal and external. Internal dynamism is the capacity of a repertoire or digital archive to change constantly, as result of interactions, changes in the people working on it, and the conditions affecting audiences and funding. The external dynamism refers to the fact that the same material can be presented in a variety of settings. For example, the same performance could be presented as a ritual and as an international entertainment in different stages, as it happens with some Balinese dance performances. In the case of a digital archive, the same content could be presented on its website, or, if the encoded XML is downloaded or served by an API, under a different interface. I use the term dynamism in the following sections, but this term could also be considered as instability or adaptability, depending on whether we want to emphasize the positive or the negative aspects of this constant change. Dynamism is not necessarily more neutral, but it is a common term and I will use it to stress both the positive and negative aspects of this propensity to change.

**Internal Dynamism**

Like performance repertoires, digital archives are constantly in flux. In the case of performances, they often change because of generational relief, changes in sociocultural conditions and in the economic forces behind sponsorship and patronage [Hughes-Freeland 2008]. For example, Balinese dances used to be primarily ritual affairs, but the dance has been modified by the influx of foreign artists and tourists. The kecak dance, now undeniably part of the repertoire learned by dancers and presented to local and foreign audiences, was partially developed by German visual artist Walter Spies [Seputtat 2012]. In digital archives, we can identify the following reasons for constant change:

- **Digital platforms encourage experimentation** and facilitate changing design decisions. Due to what Lev Manovich terms modularity, making design decisions is sometimes relatively straightforward [Manovich 2001].[2] The use of semantic styling techniques (such as SASS) for web design allows simple changes to be instantly applied to the entire website.[3] The look and feel can be constantly changed and updated and small improvements can be constantly added, or bugs fixed. However, due to the same modularity, sometimes details that would seem insignificant are indeed very difficult to correct and this can send the archives crashing. For example, changing the font and color and visual scheme of the entire website is surprisingly easy. However, as we found out working on the Asian Shakespeare Intercultural Archive (A|S|IA), ensuring a consistent integration of buttons and labels across different language scripts, such as Korean and Chinese, is a painstaking process.[4]

- **New rounds of funding.** Sometimes specific funding is available. For example, a funding body might offer a grant to develop an educational platform for an archive. Or, in order to justify a new round of funding, teams also propose fundamental changes to the design, scope, or conceptualization of an archive. Funding also means that archives are only kept for a limited time. Few archives are funded for an indeterminate amount of time, and often archival projects are dated, constrained to the duration of a grant.

- **The teams change, bringing and removing interests and expertise.** Often the archives are not impersonal, but highly modified by the interests of the participants. In the archival team where I work, the areas we work on are motivated by the expertise of the makers: Shakespearean performances in Asia, networks of theatrical production in Southeast Asia, and contemporary adaptations of wayang kulit in Indonesia.[5]

- **Practice-based approaches.** People are constantly experimenting, playing with and hacking archives. As Michelle Warren notes, digital archives are often developed through hacking, or playful experimentation
Brick and mortar archives are also in flux, but in a different way and at a slower pace. A radical change would override or erase previous iterations of a traditional archive. There is also less experimentation and “hacking” in the development of traditional archives and the disappearance of traditional archives can be more definite than that of digital ones. A digital archive can be reconstructed, if it is properly documented (more about this later, in the recommendations). A brick and mortar archive, by contrast, often disappears in dramatic ways. A tragic example of this was the destruction of the Cologne city archives, the result of a structural collapse where two people lost their lives in March 3, 2009 [Curry 2009]. It is estimated that 90% of the municipal archives of the City of Cologne, which date back to the middle ages, were destroyed by a collapse that was likely caused by the construction of a new subway station in the neighborhood. And, as a counter example, we can find the digital data stored in the hard drives destroyed by another tragic event, the 9/11 attacks. As Matthew Kirschenbaum notes, a German firm was able to restore the data from the ruins of the collapsed buildings [Kirschenbaum 2008, xii].

When traditional archives disappear, all is lost. This is also often the case for digital archives. However, in some cases, the archives merely become inaccessible or, in Tim Maly’s term, they become “dark archives.” A dark archive is a collection of materials that exists but is unavailable for general access [Maly 2013]. Crucially, though, digital archives can move in and out of this darkness, they can be kept alive and represented.

**External dynamism**

The previous section detailed the internal dynamism of archives, but we can also consider archives' potential for “external” change. Digital archives can be embedded in a different environment. This is almost impossible to do with a brick and mortar archive. But sections, or the entirety of an archive, can be reconstructed in an entirely different situation.

An item from a repertoire can easily be re-purposed for a different setting. British director Peter Brook, for instance, recounts an instance where the same performance of ta'ziyah in Iran took on completely different meanings, depending on whether it was performed at a village or at an international festival venue [Brook 1968, 16–18]. In digital archives, this capacity for the same content to be presented in different contexts is exacerbated by the quality of new media that Lev Manovich identifies as transcoding. This means that media can be transformed, converted, remixed, and inserted into other contexts [Manovich 2001, 45–47].

Following the properties of new media, digital archives can be reinserted, manipulated, and changed according to interactions with the users. There are two ways in which this external dynamism is presented:

- **Interactivity.** A digital archive can react to the user in extremely dynamic ways. Depending on how the archive is created, a user can effectively create her own sub-selection of archive materials. For example, the A|S|I|A website allows users to bookmark portions of any video and create their own annotated collection of video clips.
- **Extensibility.** Many archives also allow users to reuse the contents of the archive in another setting. The
So far, this article has explored how the conceptualization of archives in performance studies can help theorize issues that arise from digital archiving. Two new concepts, internal and external dynamism, have also been proposed to offer a more granular vocabulary to describe transience in digital archives. The next two sections look at the ways that practices from open source communities can help maintain the relevance and sustainability of digital archives. Based on the characteristics of digital archives outlined here, I have two concrete recommendations to make on how to ensure the continuity of archives, harnessing and curving their extremely unstable nature: aiming for well documented open source development, and implementing version control systems.

**Recommendation 1: Aim for (future) well documented, open source development.**

In the film *Samsara* (2011, dir. Pan Nalin), a character is puzzled by the question “how do you prevent a drop of water from drying up?” The answer, learned after long travails, is easy, beautiful, and difficultly counter-intuitive: “By throwing it into the sea.” Although there are several problems to the applicability of this philosophy to digital archiving, it does seem that opening up the source code and sharing the collection is the best way to ensure two things: the archives’ continuity in many different platforms and the possibility of “reconstructing” the archives later, in other platforms. The LOCKSS acronym – Lots of Copies Keep Stuff Safe – is well known in the computer world [D'Iorio and Barbera 2011, 66]. Ensuring that many places can legally posses copies of the source code will make the code survive. This might seem like a Faustian bargain to some, but openly sharing is the best safeguard against the disappearance of digital archives.

The reconstruction of the archive is another matter. As game studies historians will know, poorly document code is one of the main problems of trying to resurrect videogames written in languages no longer used or for platforms and consoles now obsolete. If all the source code is stashed away, there is no way of reconstructing it for later use; it is just trapped in a black box, becoming one of the dark archives mentioned by Taly. In my experience, and through conversations with colleagues, this is a difficult goal to achieve. Several problems stand in the way of ensuring open-source development.

- **Lack of understanding of open source.** Often people are insufficiently aware of the numerous licenses and processes of open source. They fear this will compromise the security or ownership claims of projects. [11]
- **Funding bodies will often not understand or allow this.** In some cases there is the pressure to develop proprietary software solutions.
- **The code is often developed by third-party companies.** It is not in their interest to share their code.

I understand it would be difficult to make things immediately accessible and shareable. But there is a workaround: make an agreement for a future disclosure of the open source. Negotiating a time acceptable to all parties, the archives’ code base could eventually become freely accessible to all. Funding bodies could also help by encouraging or even demanding this option for projects that are only funded for a few years.

An important objective for archive makers would be to also lobby or convince funding bodies of the importance of encouraging or supporting the (eventual) documentation of digital archives through open source mechanisms.

**Recommendation 2: Version Control Everything.**
How do we keep an inventory of what has gone before, of previous iterations of the archive? Often changes are made, but nobody knows about them or where they have gone. Files are overwritten or named in abstruse ways so that it is painstaking, if not impossible, to reconstruct a previous iteration of an archive. Version Control Systems (VCS) allow for creating “releases” where a particular version of the project can be fully reconstructed. This can be done in any VCS, but an obvious choice is git, which is free and often used by the open-source community.[12]

The advantage of version control is that it is inexpensive to reconstruct any previous iteration in the way that it was working earlier. This has the advantage that other people can learn from it. But also it is a reaction against the problem of constant change. It might be interesting for academic, historical, or pedagogical reasons to look back at the previous version of any given archive. Of course, both recommendations are related. It is easy to use a version control system such as git, which can be used in different platforms, to share open source projects.

CONCLUSION

Digital archives are subject to two kinds of adaptability: internal and external. In this they are similar to repertoires, which also change constantly through time and must be adapted to ever-changing circumstances. Performance studies can provide digital humanists with additional vocabulary to explore these characteristics. And the open source community can furnish us with tools and techniques in order to make the most of the repertoire qualities of digital archives.

Notes

[1] An iframe is an HTML element that allows a website to be embedded in another one. An API is an application program interface, a set of protocols that allow a program or web service to communicate with other programs or services.

[2] Modularity refers to the capacity of independent elements (layers, programming objects, etc.) to be combined into new media objects. Each elements can be independently modified and reused.

[3] SASS stands for Syntactically Awesome Style Sheets. This is a reusable, hierarchical and customizable way of defining CSS (Cascading Style Sheets) that allow for design features to be easily implemented across an entire website. See http://sass-lang.com/.

[4] The Asian Shakespeare Intercultural Archive (A|S||I|A) is an annotated collection of Shakespeare performances in East and Southeast Asia developed at the National University of Singapore. All the performance scripts, annotations, and contextual data are presented in four languages: English, Chinese, Japanese, and Korean. See http://www.a-s-i-a-web.org for more information.


[6] Responsive here refers to the practice of creating a single website that can be adapted to different screen resolutions through a grid system, as opposed to the previous practice of redirecting the user to different websites based on their screen resolution.


[8] A version of this is already available on my online dissertation website, http://www.wayangkontemporer.com. This website includes video clips that are taken directly from the CWA archive.

[9] An Indonesian language version of this is available at http://arsipbudayanusantara.or.id/.


[12] https://git-scm.com/

Works Cited


This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License.