Table of Contents for Vol. 1, No. 2
Spring 2012

Introduction

Articles

Special Section

Conversations

Archives in Context and as Context
Kate Theimer

Think Talk Make Do: Power and the Digital Humanities
Miriam Posner

Reviews

Authors

Archives in Context and as Context

KATE THEIMER
Approaching the field of digital humanities as an outsider is an interesting experience. It is best compared, I think, to being a tourist in a foreign country for which there are no reliable guidebooks. It is a country in which the language is almost the same as the one you speak, and yet words are used to mean somewhat different things. It is also a relatively young country, still trying to define its national identity.

As an archivist, attempting to learn more about this foreign country of “Digital Humanities,” I am struck with how often its citizens refer to the “archives” they or their colleagues create. To continue the tourist analogy, imagine that the country I come from is the nation of “Archives,” and that it has a longer history than that of the country of Digital Humanities. The nation of Archives has well established national principles. It is a small country, perhaps, and not a powerful player on the international stage, but its citizens are quietly proud of what they have managed to accomplish with such a small national budget.

And so I, a tourist from the country of Archives, visited the foreign land of Digital Humanities and quickly realized that something a bit odd has happened to my treasured national heritage. When I questioned digital humanists about what they meant when they use the word “archives” or questioned the appropriateness of using it to describe various collections, the responses varied from befuddled confusion (“I’m not sure what I mean”) to a strenuous...
not sure what I mean”) to a strenuous defense of the different usage. Given the emerging importance of digital humanities as a scholarly field, I thought it would be useful to explore this disconnect and so perhaps shed some light for both archivists and digital humanists about what each may mean when using this common word.

Archivists have become accustomed to the adoption of “archives” by information technologists as well as the general public to refer to things which we archivists would not call archives. So it is not the adoption of the term by digital humanists that is noteworthy, but that its meaning in certain contexts has been altered by scholars, many of whom have experience working with archives as traditionally defined. And yet it is these scholars who have chosen to describe the collections they have created as archives, seemingly in all sincerity that their usage is appropriate and not in contradiction to the practice of archivists. What could account for this disconnect?

But, perhaps more importantly, why does it matter? If some digital humanists, along with the world in general, have adopted “archives” to mean a variety of things, why should it be important to articulate and share the traditional archival vision of an archives? Archivists cannot control the use of the word “archives” and do not have exclusive rights to it. Practitioners of the digital humanities can and will continue to use it to mean whatever is meaningful in their discipline. However, I will argue that there is value and...
will argue that there is value and context in the way archives professionals have defined this term. The archivists’ definition is more specific, and therefore in my opinion conveys greater meaning. It is this meaning, and with it the understanding of the specific role archives play in preserving unique documentary material, that I want to promote.

In this article I will examine one formal definition of “archives” and use it to illustrate the fundamental principles that separate traditional archives from many of the collections created by digital humanists. I hope my discussion will itself be a demonstration of the need for greater communication between digital humanists and information professionals, such as archivists, about the areas where our practices intersect.

Surveying the landscape of the digital humanities, the “archives” that attracted my attention were primarily online groupings of digital copies of non-digital original materials, often comprised of materials (many of which are publications) located in different physical repositories or collections, purposefully selected and arranged in order to support a scholarly goal. Some prominent examples of this kind of usage are the [Shakespeare Quartos Archive](https://www.shakespearequartos.org), the [Rossetti Archive](https://www.bodleian.ox.ac.uk/rosetti) and the [William Blake Archive](https://www.blakearchive.org).[1] When I queried a few digital humanists about why they felt the collections they created qualified as archives, the most common response was that the materials had been selected. Based on this small sample, it appeared that their...
Sample, it appeared that their perception of what constituted an archive was a grouping of materials that had been purposefully selected in order to be studied and made accessible.

It is perhaps worth noting that many digital humanists, especially literary scholars, may have more direct exposure to manuscript collections or special collections, rather than true archives. The distinction between the two is sometimes not clear and many institutions have joint archives and special collections units (or departments or offices). A manuscript repository (also known as a manuscript library or special collections library) collects materials from outside sources through donation or purchase. In contrast, an archives is the repository for the historical records of its parent organization. For example, the National Archives of the United States is the repository for the historical records of the U.S. government; the Harry Ransom Center acquires its historical collections through donation or purchase. The National Archives, like most archives, also contains some donated materials; however the primary holdings of any archives will be the records of its sponsoring organization.

Although “archives” can be an organization or office within an organization, that is not, I think, the usage that is most relevant to this discussion. For that, we need to discuss the first definition of “archives” endorsed by the Society of American Archivists:

Materials created or received
by a person, family, or organization, public or private, in the conduct of their affairs and preserved because of the enduring value contained in the information they contain or as evidence of the functions and responsibilities of their creator, especially those materials maintained using the principles of provenance, original order, and collective control.[2]

There is nothing in this meaning of “archives” that references a selection activity on the part of the archivist. This led me to think that perhaps digital humanists were assuming the larger meaning of archives, which references the activities of the archivist at the repository level. This is analogous to the third definition of archives as defined by SAA:

An organization that collects the records of individuals, families, or other organizations; a collecting archives.[3]

If an archivist is perceived to be one who creates an “archives,” i.e. a place in which valuable materials are collected, then the selection function emphasized by the digital humanists makes more sense. An archivist in this sense is one who selects things for preservation and makes them accessible. And the experience of most scholars working with archival or manuscript collections may very well have left them with the
impression that this is the primary work of an archivist and the meaning of an “archives.”

And so it is, in part, but I believe that for most archivists it is the first definition of archives that distinguishes our work and our profession. Many other kinds of professionals (and non-professionals) select or collect materials, preserve them, and make them accessible.

What defines the work of an archivist, and so “an archives” in the mind of an archivist, is what materials are selected and how they are managed. Archivists select and preserve “archives” as defined in the primary definition, which is to say aggregates of materials with an organic relationship, rather than items that may be similar in some manner, but otherwise unrelated. The archival selection activity, known as “appraisal,” generally takes place at this aggregate level, and it is whole collections, donations, or records series which are being selected. These aggregates are “maintained using the principles of provenance, original order, and collective control.” These principles constitute the primary differences between archives and other kinds of collections.

The first of these principles is provenance. Just as in the art world, provenance refers to the history of an object, its creation and ownership. With works of art, provenance is usually used to better understand or authenticate an object. While those uses also apply in the archival world, provenance is also the basis for the “principle of
the basis for the "principle of provenance," also known by its French designation respect des fonds. This principle dictates that “records of different origins (provenance) be kept separate to preserve their context.”[4] In other words, records originating from different sources are never to be intermingled or combined. It is important to note in this regard that the “source” of a record is not necessarily the same as its author.

This distinction about the “source” of a record is related to the second key archival principle, that of collective control. Archival materials are generally managed as aggregates, not as collections of individual items. These aggregates, which can be referred to as record groups, series, and manuscript collections, are established according to the source of the aggregate, often a result of the activity which generated the records.[5] The principle of collective control is dependent on understanding the provenance of the aggregate of materials. To return to the primary definition of archives, the aggregate will be defined by who created it (“a person, family, or organization, public or private”) and why it was created (“in the conduct of their affairs”). The aggregate of records created by a person, family, or organization may contain records with many different authors. For example, the records of a publishing house may contain correspondence with many individual authors. Once transferred to an archival repository, those records will be maintained as a distinct aggregate (say, the “Records of Smith
Publishers”) and the contents will not be removed and added to other aggregates based on the individual authorship or topic.[6]

The third principle directs that within each aggregate of records the original order imposed by the source of records should be preserved or recreated, if it is known.[7] This principle, along with adhering to the principle of provenance and collective control, exists to preserve the original context of the records. Some records are meaningless outside their original context and others gain additional value by being examined within it.[8]

While not specified by Pearce-Moses, another defining aspect of archives is that primarily original or unique materials and not published ones are collected. When published materials or copies of materials are accessioned it is usually because they are part of an aggregate and therefore gain or provide context as part of the grouping.

These qualities taken together — preserving groups of primarily original, unique materials, which are maintained using the principles of provenance, original order, and collective control — are the bedrock of the practices of archivists. These practices are expressions of a common set of values — values which I think archivists do not discuss often enough outside our own professional communities.

I believe embedded in the discussion of what constitutes an “archives” is, consciously or not, a debate over the importance of authenticity.
importance or authenticity

the preservation of context. In fact, an essential aspect of demonstrating authenticity is preserving context. Authenticity is “typically inferred from internal and external evidence, including its physical characteristics, structure, content, and context.”[9] Physical characteristics, structure, and content are all internal evidence; the external evidence of authenticity is supplied through context, and so the archival drive to preserve context is in part motivated by the need to preserve the evidence needed to assess the authenticity of the material.

For archivists, preserving context is also about preserving the conditions that make documents more meaningful to users. All of the aspects of an archives encapsulated in the archival definition are designed to preserve the context of materials. I will return to the issue of context again, but with this in mind, I want to return to considering the digital humanities usage of “archives.”

Given the importance archivists place on the principles I have just described, it may be easier to understand the disconnect between the way archivists define “archives” and the way it is often used in the digital humanities. Archivists would not refer to online groupings of digital copies of non-digital original materials, often comprised of materials (including published materials) located in different physical repositories or collections, purposefully selected and arranged in order to support a scholarly
arranged in order to support a scholarly goal, as an "archives" — and so the confusion of an Archivist tourist in the land of Digital Humanities.

I can think of three possible responses to this archival questioning of "archives" in digital humanities. First, as noted above, archivists do select materials for acquisition and accession. So if digital humanists identify the primary activity of the archivist as one who selects things, then this could lead them to consider the collections of materials they have created by selection as "archives." However, while it is true that at the repository level, archivists create "the archives" by designating some administrative records as having permanent value and by accepting donations of collections of records created by people, families, and organizations (and occasionally purchasing them), these selection decisions are made at the aggregate level. It is these aggregates, as whole units, that are selected, not the individual items within them, which seems to contrast with the approach taken in "archives" created by digital humanists. Within an aggregate, or an "archive," archivists do not select.[10]

Second, it might be argued that the "archives" created by digital humanists are themselves archives in that they represent the records of those people’s own professional activities. For example, if digital humanist Linda Tompkins creates a digital collection of materials related to John Ruskin, do these materials not constitute "materials created or received by a person, family, organization, or institution?"
The archival response would be probably yes, but then they would be the archives of Linda Tompkins, not the John Ruskin Archives. Archivists identify aggregates, adhering to the principal of provenance, according to the source of the aggregate, not the subject.[11]

Third, it could be argued that in the digital realm a different definition of archives applies. For example, in a 2009 article in *Digital Humanities Quarterly* Kenneth Price flatly stated: “In a digital environment, *archive* has gradually come to mean a purposeful collection of surrogates.”[12] It certainly appears that this is the case in the field of digital humanities, just as information technology has adapted “archive” to mean collections of back up data. Many websites refer to the content maintained on the site, but not considered current, as existing in “archives.” All these uses are valid in their contexts. Archivists cannot control the use of the word “archives” and do not have exclusive rights to it. Language is constantly evolving and to try to enforce one group’s definition onto another group’s usage is doomed to failure. However, in such cases it is all the more important for those groups using the same word to understand the distinctions and meanings it has beyond their own borders. This is what I am trying to do here with the usages of the archival and the digital humanities.
Therefore, it is important to note that the formal definition of “archives” used in the archival community cited here recognizes no differences for electronic records, born digital material, or materials presented on the web. Price’s definition, put forward for a digital humanities audience, may be correct in that community of practice, but it should come as no surprise to digital humanists that archivists have concerns about that definition.

The issue here is not that one definition is right or wrong, but that the archival definition carries with it an adherence to professional practice and values that digital humanists are perhaps not aware of. Personally, I would prefer that online collections that do not meet the archival definition of archives be referred to as digital collections rather than archives. “Collection” clearly implies materials that have been assembled and intentionally brought together.\[13\]

However, while the purpose of an archives as traditionally defined is to preserve materials in their original context (or at least “the organizational, functional, and operational circumstances surrounding materials’ creation, receipt, storage, or use, and its relationship to other materials”\[14\]), archivists recognize that this is by no means the only context in which materials may be understood. For example, a letter written by Dante Gabriel Rossetti may have context within the records of an art dealer or
within the records of an art dealer or publisher preserved in an archives, but it will also have context seen with his other correspondence as gathered together in the online collection that is the “Rossetti Archive.” The critical difference is that while such a letter can be placed within many different contexts in many different kinds of collections, it is only in a collection managed according to archival principles that the organizational context of the letter is preserved. Preservation of this kind of context is what separates archives from libraries, most personal collections, and assembled virtual collections.

What concerns me is that in the broadening of “archives” to extend to any digital collection of surrogates there is the potential for a loss of understanding and appreciation of the historical context that archives preserve in their collections, and the unique role that archives play as custodians of materials in this context. Given the connotations of authority, rarity, and “specialness” that the word “archives” has in our culture, it is not surprising that it is an attractive word to use, as the creators of the William Blake Archive admit, to describe an online collection for which no other word seems to fit. I have no illusions that this discussion will alter how digital humanities scholars use “archives” within their own projects and discourse. I do hope, however, that this usage can be informed with an understanding of the principles embedded in the word as archivists have defined it, and that the role of archives (the kind that archivists
archives (the kind that archivists manage) as custodians of a particular kind of context can be appreciated.

Expanded from an original post by Kate Theimer on March 27, 2012. Revised for the Journal of Digital Humanities June 2012.

[1] Interestingly, the William Blake Archive provides an explanation of “What do we mean by an ‘Archive?’” which concludes with “Though ‘archive’ is the term we have fallen back on, in fact we envision a unique resource unlike any other currently available for the study of Blake—a hybrid all-in-one edition, catalogue, database, and set of scholarly tools capable of taking full advantage of the opportunities offered by new information technology.” I read this as confirming that they knowingly used the word “archive” to describe something that they knew was not actually an archive since they describe it as a “unique resource.”


[3] Ibid.


[5] For definitions of these terms, refer
to Pearce-Moses, *A Glossary of Archives and Records Terminology*. Note that while “manuscript collection” might seem to refer to a collection of manuscripts, it is used specifically to describe a collection of personal or family papers. The use of the word “collection” in archival practice can be confusing. It can also be used in the sense of an “artificial collection” which is “A collection of materials with different provenance assembled and organized to facilitate its management or use.”

[6] Perhaps the most prominent feature of collective control is that archival collections are described as aggregates (again, as record groups, collections, and series) but rarely, if ever are the individual items in an aggregate described. This difference in the level and type of management and description is often cited as a key differentiation between archives and libraries.

[7] As Pearce-Moses notes in his definition of “original order”: “A collection may not have meaningful order if the creator stored items in a haphazard fashion. In such instances, archivists often impose order on the materials to facilitate arrangement and description. The principle of respect for original order does not extend to respect for original chaos.”

[8] For a classic case of the value of context, see the example summarized in “Well done': When context of records matters.”

[10] Some selection occurs below the aggregate level in the activity commonly referred to as “weeding” (or “culling”). Weeding may occur during the processing of the collection, and refers to the removal of material deemed to have no value. Examples of materials which may be weeded are duplicate copies of materials, blank letterhead or stationery, etc.


[13] That said, it should be noted that some information professionals are adopting the usage of the digital humanities community by referring to their own assembled collections of digital copies as “digital archives.” See for example the Marcel Breuer Digital Archive.


About Kate Theimer
Kate Theimer is the author of the popular blog and Twitter account ArchivesNext and a frequent writer, speaker, and commentator on issues related to the future of archives. She is the editor of the series *Innovative Practices in Archives and Special Collections*, in which books on description, management, outreach, and reference and access were published in 2014. She is the author of *Web 2.0 Tools and Strategies for Archives and Local History Collections* and the editor of *A Different Kind of Web: New Connections between Archives and Our Users*. She has published articles in *The American Archivist* and the *Journal of Digital Humanities*. Kate served on the Council of the Society of American Archivists from 2010 to 2013. Before starting her career as an independent writer and editor, she worked in the policy division of the National Archives and Records Administration in College Park, Maryland. She holds an MSI with a specialization in archives and records management from the University of Michigan and an MA in art history from the University of Maryland.