11-1-2012

Context and Its Role in the Digital Preservation of Cultural Objects

Joan E. Beaudoin
Wayne State University, joan.beaudoin@wayne.edu

Context and Its Role in the Digital Preservation of Cultural Objects

Joan E. Beaudoin, Wayne State University

joan.beaudoin@wayne.edu

Abstract   In discussions surrounding digital preservation, context — those properties of an object related to its creation and preservation that make the object's origins, composition, and purpose clear — has been identified as a critical aspect of preservation metadata. Understanding a cultural object's context, in as much detail as possible, is necessary to the successful future use of that object, regardless of its form. The necessity of capturing data about the creation of digital resources and the technical details of the preservation process, has generally been agreed. Capturing many other contextual aspects — such as utility, history, curation, authenticity — that would certainly contribute to successful retrieval, assessment, management, access, and use of preserved digital content, has not been adequately addressed or codified. Recording these aspects of contextual information is especially important for physical objects that are digitally preserved, and thereby removed from their original setting. This paper investigates the various discussions in the literature surrounding contextual information, and then presents a framework which makes explicit the various dimensions of context which have been identified as useful for digital preservation efforts, and offers a way to ensure the capture those aspects of an object's context that are often missed.

Introduction

> "The context of a digital object to be preserved over time comprises the representation of all known properties associated with it and of all operations that have been carried out on it." (Brocks et al., 2009, p. 197)

This paper seeks to examine and clarify contextual information recorded for the preservation of digital cultural objects. An overview of the

published literature written on the topic of contextual information recorded for digital preservation is provided here to illustrate the multifarious nature of the topic. The various approaches to the topic of context revealed through the literature are then used to develop a multidimensional framework within which to capture contextual information regarding cultural objects. This framework acknowledges the rich information about context that can be captured to provide more effective means of search, retrieval, examination, use, management, and preservation for cultural objects in a digital form.

Digital preservation, according to Conway (1996) is the "acquisition, organization, and distribution of resources to prevent further deterioration or renew the usability of selected groups of materials." This definition provides an indication of the various efforts involved in preserving digital materials so that they find extended use, but it leaves a key piece of the preservation process unacknowledged. The importance of preserving the descriptive and explanatory information that accompanies digitized materials fails to appear in this definition, except perhaps through intimation. This situation is not surprising given that preserving digital content is the principal goal of digital preservation. The literature surrounding digital preservation reflects this aim, and so it has primarily focused on those technical issues that need to be addressed in order to extend the life of digital materials beyond their period of creation. However, this focus means that the important contextual data concerning digital content generally go unrecognized. This situation exacerbates the contextual break that occurs in the information available about an item beyond the time of its creation. The further removed the period of creation of an object (digital or otherwise) is from the period of its later examination, the less likely it is that its full significance will be appreciated. Knowledge about the context of cultural objects is nearly mandatory for our understanding, use, care, and preservation of them. An acknowledgement of this situation can be seen in the investigations of several researchers who have considered issues of contextual information for digital preservation.

Many authors have discussed the general problems encountered when there is a lack of contextual information. One of the earliest authors to address this problem in the literature felt that the predominantly technical metadata recorded at the time of a digital object's creation was of

limited usefulness since it lacked information concerning the historical context, or broader contextual information beyond that of the current system (Duranti, 1995). Even at this early date in the discussion of digital preservation, the limitations of information recorded during the digitization phase were recognized. This focus on the technical details has remained a common topic in the literature in the intervening years. Chowdhury (2010) noted that the primary topics addressed in the digital preservation literature are those which focus on technological and semantic information surrounding digital content. While technical details are useful in their own right for the preservation record of digital objects, this does little to aid our broader understanding of the item. The difficulties resulting from a restricted view of context in digital preservation metadata appear in more recent discussions of the topic, with several authors expanding the discussion to include very different kinds of metadata (Lavoie & Gartner, 2005; Watry, 2007; Lee, 2011).

Several authors have discussed the need and reason for recording contextual metadata. Conway (1996) notes the difficulties encountered with a lack of contextual information for digital materials, stating that this creates a situation where "... we find ourselves confronting a dilemma such as the one faced by Howard Carson, Macaulay's amateur digger [in *Motel of the Mysteries* (1979)]: a vast void of knowledge filled by myth and speculation." For Conway, preservation is primarily concerned with evidence that is a part of the physical object and the intellectual content represented by it. Digital materials for him, since they are divorced from the physical world, are seen as fragile objects in perpetual danger of loss or damage without the information needed to contextualize them. Lee (2011) also uses an archaeological analogy in his paper examining the topic of contextual information within digital preservation, noting that the difference between an archaeologist and a looter is that the latter does not record contextual information before removing objects from their find spot. Removing an object from its surrounding stratigraphy without recording those details often means that interpretive clues and the object's full significance are lost. While most authors would now recognize that there are multiple levels of contextual information useful to digital preservation, the problem may be the lack of resources available to the task. Watry (2007), in fact, questions whether sufficient capture and management of contextual metadata are achievable for meeting the needs of the archivist and, I would add, the ultimate users of preserved digital

content.

Owing to the relatively youthful nature of the discipline of digital preservation, with its limited exploration and tentative practices, a marked tendency toward addressing fundamental principles has appeared in the literature. This can be seen in Bearman's (2007) discussion of digital preservation where he notes there is little consensus about fundamental issues of what should be saved or how to save it. This idea of worthiness is mirrored by Vogt-O'Connor (2000) when she suggests criteria to be used in choosing materials for digitization projects. The evaluative questions she asks concerning selection indicate the critical nature of context in the digitization process. She asks "[d]oes the candidate material require substantial research and a sophisticated and expensive context in order to be useful?" (Vogt-O'Connor, 2000, p. 68). Indicating just how critical this information can be for their use, she goes on to state that if context for the materials being digitized cannot be provided, other materials should be chosen. Expanding upon these selection rules for the digitization process itself, it seems likely that these criteria should also be employed in decisions concerning digital preservation efforts.

One of the most difficult problems encountered in the discussion of context as it relates to digital material is the variable nature of the term. Vogt-O'Connor used the term in the discussion above to express possible technical limitations of the digital materials themselves (or their systems) which would interfere with the reception of key characteristics of the physical objects. However, the meaning of the term context in the passage above could just as easily be applied to discussions about social, historical, physical, or a whole host of other aspects. It was only through a reading of the text surrounding the above passage that the specific meaning of context was discovered. The text served as the "contextualizer" for the term in this instance. This discussion concerning Vogt-O'Connor's passage offers a brief, but clear example of how important context is for the reception of information. The problems of context can be exacerbated in the case of non-textual media, such as visual or audio materials, as they often do not include text to provide contextual clues.

Context is especially important in discussions of digital preservation since in most instances the digital materials have been separated from their original format and context in the processes of digitization and preservation. Digital materials pose a "... risk of decontextualization — the

possibility that the digital surrogate will become detached from some context that is important to understanding what it is, and will be received and understood in the absence of that context", (Unsworth, 2004). In other words, since digital materials are typically not situated within their original context they are prone to being experienced and interpreted in ways that were unintended. While there is value in using materials in decontextualized ways, for example, as a sort of creative springboard, it is critical that the original and intended meaning and/or experience be preserved whenever possible.

Contextual information surrounding digital content is varied. What follows is a discussion of eight major preservation topic areas that were identified during a review of the digital preservation literature that addresses the concept of context.

Technological Aspect

By far the most thoroughly investigated form of context in the literature surrounding digital preservation is that concerned with technology. As was mentioned earlier, this is hardly surprising given the centrality of this topic to the discipline of digital preservation. Issues of hardware and software, emulation and migration, formatting, and translation all fall under this general rubric and are issues that continue to receive much research interest. Day (1997) is among the earliest authors to discuss the importance of recording technological context for digital preservation. He suggested that Dublin Core elements could be used to preserve details (e.g., migration, encoding) about the technical context of digital materials. Furthermore Day (1997) suggests that the metadata recorded for each instance would make it possible to discover how to accurately manipulate and display digital materials.

Discussions of the issues surrounding technical context can be found in the work of Levy (1998), Bullock (1999), Besser (2000), and Chen (2001). Beyond the technical dependencies of digital materials on hardware and software, these authors address technological issues such as emulation, file formats, migration, storage, obsolete hardware maintenance, compression and encryption and how these have important implications for the future reuse of preserved digital content. Bullock (1999), Levy (1998) and Chen (2001) discuss the difficulties facing any preservation

effort due to the history of rapid obsolescence and lack of backward compatibility found in the digital arena. Chen (2001) suggests there are diametrically opposed needs in the area of digital materials. This is seen in the need to maintain digital materials intact as they were created, while at the same time wanting to use ever more advanced tools and techniques. Levy (1998), too, argues that there is a division between the technical requirements of digital preservation and the users of those materials, and so he states that "[t]he challenge ahead is to bring our best technical skills to bear on the problem of digital preservation without losing sight of the ultimate human purposes these efforts serve, purposes which cannot be found within machines", (p. 161). For Chen (2001) the disparity between how digital context was created and how it was used represents a major research challenge, as well as requiring increasing amounts of metadata.

The importance of metadata to record technical information for digital preservation, mentioned by Day (1997) and Chen (2001) is more completely addressed by Waibel (2003), Brocks *et al.* (2009), and Faniel & Yakel (2011). Waibel (2003) discusses the topic of technical context through three interlocking metadata standards, the Open Archival Information System (OAIS), Metadata Encoding and Transmission Standard (METS) and NISO Data Dictionary — Technical Metadata for Digital Still Images (X39.87). Using these, Waibel attempts to capture the full spectrum of information surrounding the preservation of digital materials. Technical aspects of context were similarly the focus of Brocks *et al.* (2009) in their paper which developed an extended OAIS model for digital preservation. Digital preservation is not just a technical problem, however, as Chen and Levy observed. For digital preservation to be successful additional aspects beyond technical details need to be recorded for digital content.

A broadening of the kinds of information to be recorded is evident in the paper by Faniel & Yakel (2011) where they state that "[c]ontextual metadata hasn't garnered a great deal of attention, but there is an acknowledgement that it is key to long-term renderability and meaningfulness in reuse", (p. 156). These authors go on to state there are currently two separate research camps, that of digital curation and that of reuse, and that these two camps focus on different aspects of preservation metadata. The digital curation camp focuses its attention on metadata for technical aspects in digital preservation, while the reuse camp examines

meaning making through metadata. Recording multiple kinds of context about digital content is also a topic addressed by Mayer & Rauber (2009) in their paper which introduces semi-automatic methods to capture information critical to the interpretation, authenticity and use of large data sets. Using the dimensions of time, object type, contributors and content these authors examine how contextual information can be detected and extracted from digital objects embedded in an information space. While technical details have been a primary focus of discussions surrounding digital preservation, the future utility of the preserved items is an often identified reason for including contextual data and so this topic is what we turn to next.

Utilization Aspect

Context in this case clarifies aspects about who the audience is and what their requirements are when they seek out and use digital materials. The importance of use context is seen in Hedstrom's (1998) definition of digital preservation "... as those methods and technologies necessary to ensure digital information of continuing value remains accessible", (p. 190). In order for digital materials to remain accessible, preservation efforts must ensure that the requirements of users, present and future, are met. Wallis *et al.* (2008), in their study of eScience data archiving and reuse, discuss how the quality and value of digital content are tied to a user's ability to understand its origins, provenance, and context. Particularly important to these researchers was the documentation of decisions on what content was retained and how it had been processed (collected, cleaned, calibrated, reduced, etc.) prior to its original use and deposition in the digital archive. While these researchers examined eScience data rather than cultural heritage objects, their study helps point out that digital content may pass through various stages of use and reuse. As circumstances of use have been recognized as crucial to a determination of what is to be preserved, recording contextual information about use would be helpful (Levy, 1998).

There is, however, some disagreement among researchers about how important users ultimately are in the digital setting and what aspects of use, including the needs of the users themselves and their specific tasks, required tools and social, political and/or organizational settings, should be considered. The degree to which potential users and uses of an object

can be judged with any accuracy has been debated by Lynch (2002), who states that "... perhaps we should avoid over-emphasizing pre-conceived notions about user communities when creating digital collection[s] at least in part because we are so bad at identifying or predicting these target communities." While it may be difficult to predict who the eventual users of digital objects may be, it is fairly clear that the impetus to digitize materials or provide access to born-digital content typically originates with some defined audience in mind. Marchionini & Maurer (1995) identify three basic types of users of digital materials in an online setting. While specifically written for an audience interested in digital materials for educational purposes, these authors outline the various types of "learning" experienced by users of digital libraries and offer a discussion of the levels of intermediation needed by each. They suggest that the creation of an intellectual infrastructure for the effective use of materials is dependent on the user type (formal, informal, or professional).

A categorization of digital content users into types (expert, general, or casual) is also discussed by Benoit (2011) in his study of how information systems which contained contextualizing information about items held in it were perceived by various groups. Benoit's study is useful to note here since it offers support for the idea that contextual information about use plays an important role in information seeking. Users without specialized subject knowledge, those falling in Benoit's general user classification, "felt they could pose a broader range of (unusual) questions that are meaningful to their information needs", (p. 144). Furthermore, Benoit found that the "integration of user context-use data altered expectations of the role of information systems in general", (p. 144). In addition to the benefits suggested for the ultimate end-users of preserved digital content, Copeland & Barreau (2011) note that user-supplied contextualizing information may assist people in identifying, preserving and sharing their own digital content.

Aspects of use incorporated into retrieval systems ensure the future value and usefulness of digital materials and so they should be recorded. Specific task-based needs of users can be all-important in the use of digital materials, as Meyyappan *et al.* (2001) and Mayer & Rauber (2009) discuss. Digital preservation must also consider the tools and techniques used to support users' analyses. For example, in a scholarly setting, tools to help with interpretive processes, note taking and collaboration have been noted

as important aspects of use (Palmer, 2002). Mayer & Rauber (2009) present several use scenarios where automatically generated contextual information is used to assist "in virtually any task where specific digital objects are concerned and where the context is not obvious to the user", (p. 8). While digital materials are dependent on the systems and tools developed for their presentation and usage, they can become separated from their mechanisms of presentation and usage and so some provision must be made to identify how the materials were intended to be used by their primary audience.

A critical aspect of use to be discussed in the context of digital preservation is the original setting for the digital materials. Social, political and/or organizational contexts have a broad impact upon the use of digital materials and these aspects should also be recorded in the preservation record. As Adams & Blandford (2004) discovered with their study of digital libraries within a medical setting, the use of digital materials cannot be divorced from a critical analysis of the social and organizational setting within which their users operate. These researchers found that inadequate consideration of these aspects can lead to negative perceptions of digital libraries, a lack of knowledge about, abilities with, and awareness of digital libraries, and can contribute to the hoarding of information and technology. As users are so important to the use and reuse of digital materials, aspects concerning the intended use and audience also need to be addressed through the metadata record for digital preservation.

Physical Aspect

Many of the difficulties experienced with digital preservation are simply due to the fact that digital materials are decontextualized from their original state in the digitization process. Simple characteristics of the original are lost in the creation of a digital surrogate of that work. Information about scale, surface, behavior, relationships, arrangement of parts, functionality and so on, is intimately tied to the perception of physical objects. Digital materials, while they enable some analyses which are impossible with physical manifestations, provide very weak information concerning tangible aspects. Bullock (1999) states the theme of documentation and description in the digital realm is in part due to the fact that digital objects tend not to carry visible evidence of their creation.

Clues to information concerning the original objects, such as those found in the materials and techniques used in their creation, tend not to be readily discernible in digital surrogates. While physical aspects are fundamental to the reception of the digital object in its use environment, they also guide preservation decisions. Without information concerning the physical nature of the original it is difficult to make informed decisions about which digital items should be selected for preservation efforts.

Another aspect that has been discussed concerns how user experiences differ between the original and digital versions. As Meirelles (2004) points out in her paper on the challenges of presenting artworks in the electronic environment, the way an item is experienced is mediated through hardware and software. Visual displays, speakers, system speeds, interface design, mice and other devices used to interact with digital content transform how the original is received. That changes in an item's reception can occur due to hardware and software variations, even with objects created for the electronic environment, speaks to the basic problems inherent in the medium.

Issues with the physical-digital transformation are apparent in the discussion of decontextualized digital materials by Unsworth (2004) and Conway (2009). Conway (2009) carefully recounts how the digitization of historical photographs "diminishes, masks, or even distorts visual cues that are potentially fundamental to the extraction of meaning", (p. 16). The relationship between representation, replacement, and superiority in the physical-digital transformation are complex and fraught with many challenges. Due to these problematic relationships, Menne-Haritz and Brübach (n.d.) feel that through the conversion process critical information about the context circumstances of documents/objects is lost, and so "[d]igital imaging is not suitable for permanent storage." These authors suggest that since digital materials are unable to accurately represent analog objects, there is little reason to be concerned with digital preservation. Unsworth's (2004) suggestion that each digital surrogate is "shaped by the perspective from which it was produced", also alludes to the limitations of digital materials to truthfully represent original objects. The result of the analog transition to digital media is multiple and varied versions of a single item. The question of how we choose the one that most closely reflects the original remains unanswered. Conway (2009), in

his discussion of ways to regulate or lessen the loss of information in the analog to digital transformation, points to the potential usefulness that standardized digitization guidelines and explicit processing statements could provide.

A number of the problems experienced in the physical-digital transformation are due to the fact that, unlike physical materials, formats and principles for digital preservation are in the early stages of development. Problems associated with the lack of persistency, how digital objects relate to one another, the behavior of digital objects, and so forth, could potentially be resolved in the long-term when fully developed methods and principles are available (Besser, 2000). On the other hand, there may be viable reasons to represent materials in their original, historical format. Without the ability to provide an object's original access and functionality, the experience of the user-viewer no longer reflects what was intended by the item's creator. In this case, the ability to record what is to be retained, perhaps through a statement of the creator's intentions, is of paramount importance in guiding preservation efforts (Lusenet, 2002).

Intangible Aspect

Although typically not mentioned outside of discussions of the physical features lost in digitization of items, this dimension of context is concerned with recording those aspects which are the result of the intangible nature of digital materials, and so is an aspect believed to warrant its own entry. This aspect includes qualities such as indistinct object boundaries and impermanent linkages between digital materials. Meirelles (2004) notes that interactions, links and connections made between data are important to the way a work is experienced. This suggests that the vague and sometimes shifting nature of digital items, as is discussed by Besser (2000), Bullock (1999) and Lusenet (2002), has a powerful influence on how we receive digital content.

Curatorial Aspect

Although this aspect hasn't received much attention in the literature, several authors have mentioned issues surrounding the custodial tradition of the information record of digital materials (Gilliand-Swetland, 2000;

Lavoie & Dempsey, 2004). This aspect is concerned with the care and protection of digital content, and the preservation of the information surrounding these objects. Besser (2000) suggests that digital preservation efforts have been stymied due to the fact that issues of responsibility between librarians and technical staff have yet to be resolved. Besser suggests that if neither group claims responsibility for this effort, it is likely that this work will never be carried out in any systematic way. While Nesmith (2005) discusses context as it relates to the construction of records within the archive, he suggests that the custodial history, the use of archival materials, and the impact of records across time can be used to "... explain why the records exist, what they might be useful evidence of, and how they have been and might be used", (p. 271). Thus, in providing information about the custodial history in the preservation record, future users will be privy to the reasons relating to why the digital objects exists and the decisions that were made for their preservation.

Authentication Aspect

Authentication context, those issues of digital preservation surrounding evidence and verification, has garnered a great deal of attention in the literature surrounding archival records. Hedstrom (1998) notes that the ability to judge and authenticate the integrity of a source is particularly problematic with digital materials since they are so "... easily altered, copied and removed from their original context", (p. 192). Gilliland-Swetland (2000) also notes the difficulties of amassing evidence with materials that show little chain of custody. One way to authenticate these materials is to "... require archives and libraries to preserve contextual and descriptive information", in addition to the original content (Hedstrom, 1998, p. 192).

More recently Duranti (2005) states, while writing on the topic of the long-term preservation of digital records, that in order to preserve authenticity of the records, the identity and integrity of the content must be maintained. She suggests that the identity of digital content can readily be maintained through metadata directly attached to the material being described. Integrity, however, presents several challenges. Difficulties associated with verifying the integrity of digital content can result from the proprietary nature of specific environments within which the materials reside. According to Duranti (2010), one way to alleviate this

problematic situation is through the use of open source environments as they are able to satisfy the "legal requirements of objectivity, transparency, verifiability and repeatability for any process that is carried out in a digital environment", (p. 163). Mayer & Rauber (2009) state that advanced tools, such as automatically generated contextual analyses, can help to eliminate the difficulties encountered in the tasks associated with manually identifying and establishing the provenance of the digital content. Although a high level of interest in the authentication of digital content has not been reflected in the literature surrounding cultural materials, archival investigations into issues such as provenance, tracking content changes, integrity, and versioning are likely to be equally applicable in the sphere of cultural heritage.

Authorization Aspect

Information concerning the intellectual property rights of original objects and their digital surrogates is another topic that was found in the literature. Aspects which fall under this type of context include information concerning rights holder(s), rights management, and allowable legal use. Surprisingly, discussions of intellectual property rights within the realm of digital preservation for cultural heritage literature are uncommon. The rights of original content producers are, however, addressed within the cultural heritage community and this topic also appears in studies that examine the importance of documentation of ownership of digital content (Ormond-Parker & Sloggett, 2011). Lavoie & Dempsey (2004) offer a brief discussion of issues surrounding intellectual property rights in the realm of digital preservation. These authors suggest that intellectual property rights for digital materials are ambiguous under the current law, and that there are two competing issues at play in the preservation of digital content: the need to intervene to preserve digital materials and the need to protect intellectual property rights. Besek (2003) and Hirtle *et al*. (2009) present overviews of the rights, exceptions and responsibilities associated with copyright and digital materials that are generally applicable for cultural objects.

Digital preservation is an item of concern to copyright holders since its processes require copying, and in some cases migration, of content in ways that change the original digital object. Duranti (2010) discusses these issues in the context of digital preservation and states that the intellectual

property rights of the copyright holder are coupled with the authentic version of the digital content. Transformative migration is particularly important to preservation efforts, according to Duranti (2010), as "additions or modifications to an existing work ... can trigger new copyright considerations", (p. 160 n. 3). As digital content has an additional level of complexity concerning intellectual property rights, metadata to record aspects to lessen future challenges would likely be welcomed.

Intellectual Aspect

One category of context which has a strong tradition in the scholarship associated with the cultural heritage community is information surrounding the significance of cultural objects. This category of context includes aspects such as meaning, function, technique, historical importance, narratives and communication of ideas through cultural objects and, by proxy, their digital counterparts. Understanding a digital object's original intellectual context is viewed as critical to the reception of a work by a number of authors writing in the service of archives, libraries and information science (Bullock, 1999; Besser, 2000; Lusenet, 2002; Lynch, 2002; Dalbello, 2004; Mayer & Rauber, 2009; Duff *et al.*, 2011; Wisser, 2011). These authors note that basic questions about meaning, function, presentation and orientation can be answered through information recorded to contextualize objects in meaningful ways. While recording this form of information is noted as being critical to the future interpretation and use of preserved digital content, a basic framework to capture contextual information to assist in the future understanding of the intellectual milieu of digital content has yet to be codified and adopted among the cultural heritage community.

While there is a deeply rooted tradition of recording information concerning materials in the cultural heritage community, McCarthy (2007) notes that the management of this information has been difficult to put into practice. According to McCarthy, the inadequate preservation of digital content leads to an epistemic failure, a lack of information required for an understanding of the structure and meaning of the metadata. Although speaking from a place where contextual information is

envisioned more broadly than only that concerned with the intellectual aspects of digital content, McCarthy (2007) directly addresses the critical nature of this information by stating that "the present generation, with its knowledge of the resources, has a clear obligation to preserve that knowledge and pass it on to future curators so informed decisions on future management can be made", (p. 256). Because it plays such a critical role in future understanding, contextual information surrounding digital content needs to be seen as an integral component and not merely optional data to be captured when time and funding allow for it.

The importance of metadata to the future understanding of the intellectual aspects of preserved digital content has been an often discussed topic in the digital preservation literature. The 2009 draft OAIS standards, produced by the Consultative Committee for Space Data Systems, draws attention to the fact that descriptive information about digital content is needed to maximize future use and understanding of preserved objects. Descriptive information about the digital content being preserved appears in several critical areas of the OAIS model, and in fact the model contains an area titled "Preservation Description Information (PDI)", specifically to record information for preservation purposes. The PDI area focuses on "information that will support the trust in, the access to and context of the Content Information over an indefinite period of time", (CCSDS, 2009, p. 4-28). Aspects to be included in the PDI consist of information concerning reference resources, context of creation, origins and provenance, data integrity (fixity) and rights. As useful as the OAIS model is for identifying the kinds of information to be recorded, it is meant to be broadly useful in a variety of settings. Thus, its coverage of descriptive information is general in nature and does not include a proscribed metadata schema for capturing this information.

A framework for contextual information of a primarily intellectual nature for digital content is presented by Lee in his 2011 article titled, "A framework for contextual information in digital collections." Contextual information for his purpose falls into three specific areas: the formation of meaning, the situation of the object and the situation of the user. The first form has to do with the formation of meaning via the surrounding environment (e.g., meaning of a word embedded in a passage). The second form has to do with characteristics or conditions surrounding the object (e.g., location, social setting, or placement). The final form has to do

with the situation or state of the user which influences interpretation or understanding (e.g., priming, situational relevance). Using this as the basis for his later discussion Lee (2011) goes on to develop a framework with nine classes of contextual entities that he believes are particularly useful to capturing information useful to the intellectual aspects surrounding digital content. These nine classes are identified as object, agent, occurrence, purpose, time, place, form of expression, concept or abstraction and relationship (Lee, 2011, Table I, p. 106).

Several current research interests in parallel disciplines could also contribute to the development of a metadata schema to record intellectual context. For example, interest in developing metadata schemas for contextual information about research data sets in the scientific disciplines could be examined for aspects that would also apply to cultural materials. Cowan & Lillico (2009) present a metadata framework for recording information about research projects where they include the project's title, individuals on the project team, funding organizations, account and file codes, dates, status, summary, publications, files, where data was gathered from, who gathered the data, and when the data was collected (Table 6, p. 99-100). Also addressing issues of research data, Wallis *et al*. (2008) discuss how critical the interpretative metadata is to researchers, since they often have little knowledge of who has acted upon the data or what has been done to it. These authors present a nine stage life-cycle model which identifies the various processes the research data may pass through during each of these stages. These processes provide critical points to be highlighted in the documentation of each stage which would be useful to later understanding of the data.

Various methods of capturing contextual information are currently available. The most commonly encountered method for representing cultural objects within their intellectual contexts are human-mediated descriptive accounts. Providing this form of context is an important step which allows future users to experience or understand the item as it was originally intended. Richer modes of documentation are available, however, as Carrozzino *et al*. (2010) point out in their article, which examines a 3D virtual interactive platform to try to capture long-held bronze casting skills important to the culture and history of Lucchesia, Italy that are being lost. Other modes of capturing the intellectual context of digital content are the semi-automatic methods described by Mayer &

Rauber (2009). These authors describe how visualizations of, and interactions with, large bodies of digital content can reduce the manual work involved in traditional methods of capturing intellectual context. Although this captured information is limited to the context surrounding digital content rather than analog objects, semi-automatic methods were shown to capture information surrounding when an item was created and the individuals associated with the content. Both of these are standard aspects of the intellectual context of cultural materials. While descriptive information about digital content isn't necessarily critical for its use, it does add important details to what has been recognized as an imperfect representation of analog content.

The scholarship surrounding the documentation of intellectual context has been strongly influenced by postmodern theory. Postmodern theory posits that all acts of description and interpretation are influenced by circumstances surrounding the author and this in turn creates a fragmentary and ever-shifting view of truth. Thus, all descriptions and interpretations are limited in their ability to fully explain the truth about cultural objects. If we accept the postmodern stance about the permutable nature of truth, should information about context be recorded at all? A number of scholars suggest there is no such thing as a neutral interpretation of cultural materials, and yet they support efforts to continue recording information about materials (Buckland, 1988; Lynch, 2002; Nesmith, 2005; Duff *et al*., 2011). In fact, Nesmith (2005) feels the contextualization of materials is an ongoing process and states that "... more context is always needed if we are to understand what is possible to know", (p. 260). For these authors the act of interpreting an object has value in that it adds an additional layer of information about a work, and interpretations should appear as a part of the work's intellectual record.

A related development is the marked focus on the interpretation of materials in the literature surrounding digital libraries and preservation. The first of these was Bénel *et al.*'s (2001) article discussing the interpretive description, based on an idea of truth that is situated firmly within social, historical, cultural and action related contexts. According to these authors this approach supports the positive goals of communication, collaborative use of vocabulary and sense-building across a group. This interest in interpretation can also be seen as a call for developing interactions within a digital library setting which present a richer user experience than the

typical functions found in current online collections.

Dalbello's (2004) study of digital libraries is also useful to consider here as she found a preponderance of presentation techniques for materials which offered "disengaged objects in search of narrative coherence", (p. 282). Since digital materials are generally presented in systems providing a display-focused experience to the user, Dalbello found a lack of comprehensiveness and closure. What was missing from the users' experiences with the digital libraries, according to Darbello, was "contextual processing." Similarly, Lynch (2002) finds that digitized collections of cultural materials are in need of additional work to package the content in ways that foster users' learning experiences, interpretations and analyses. Because of these efforts, Lynch (2002) notes that the historically separate roles of librarian, scholar, curator and teacher are blurring alongside the traditional distinctions between libraries, museums and archives. Extending this idea a bit further, many authors on this subject note the importance of community interaction with, and interpretation of, digital materials (Bénel*et al.*, 2001; Lynch, 2002; Dalbello, 2004; Unsworth, 2004; Lagoze *et al.*, 2005). Lagoze *et al.* (2005) sum up the others' ideas stating "[t]his added value consists of establishing context around those resources, enriching them with new information and relationships that express the usage patterns and knowledge of the library community. The digital library then becomes a context for information collaboration and accumulation — much more than just a place to find information and access it."

This idea of accumulating layers of information around digital materials through interactions with and responses to content is one that echoes the words of Brown & Duguid (1996) in their seminal article, "The social life of documents". Cultural materials, like text-based documents, acquire rich intellectual substance over time. Unfortunately, unlike text-based conversations which can be traced through citation records, connections between the various intellectual exchanges surrounding cultural materials are more tenuous. This is a critical reason to support the documentation of contextual information, although not the only benefit to be had for the development of a framework to record this information. McCarthy (2007) discusses the various benefits of recording information about digital content and suggests that these include support for knowledge transfer, decision-making processes, improving transparency

(and thus, build trust), providing a structured and visible system for knowledge sources, and "vastly improving discovery, accessibility, and comprehensibility of resources", (p. 254). It was with these benefits in mind that the current research was undertaken.

The Dimensions of Context

This examination of the literature was conducted to identify the important dimensions of context and how they apply to the preservation of digital objects, and to aid in the development of a framework for recording contextual information. Eight distinct dimensions of context, which make explicit the various forms of context identified as useful to digital preservation in the literature, are presented in Table 1 below. Each dimension has multiple characteristics which are further developed, along with the framework itself, in the second phase of this work described in a paper also published in *D-Lib Magazine*[1].

Table 1: Dimensions of context.

Technical: This dimension of context concerns digitization processes and techniques. This includes aspects such as file formats, hardware, software, operating systems, migration, emulation, storage, data loss, encapsulation of technical information, and compatibility.

Utilization: This dimension of context speaks to the needs of users. It includes audience needs, task support, tools, accessibility, audience characteristics, and the types of analyses to be supported.

Physical: This dimension of context speaks to those characteristics of a work that are dependent on a direct, tangible interaction. This includes features of analog and digital items which are sensory in nature, and so includes all issues relating to the object's physical presence (e.g., scale, materials, texture, arrangement, sound, brightness, smell, etc.).

Intangible: This dimension of context concerns the intangible nature of digital materials. This includes qualities such as indistinct object boundaries, impermanent relationships and network linkages between digital items.

Curatorial: This dimension of context is related to the standards and guidelines used in the preservation process. This includes facets such as the tradition of stewardship, and preservation purposes and strategies.

Authentication: This dimension of context is connected to evidence and verification. This includes the provenance, tracking of content changes, integrity, and versioning that occurs with digital items.

Authorization: This dimension of context concerns the intellectual property rights surrounding the original object and its digital surrogate(s). This includes aspects such as rights management, legal usage, and rights holder(s).

Intellectual: This dimension of context is concerned with the significance of the original cultural object and, by proxy, its digital surrogate(s). This includes facets such as meaning, function, creative technique, historical import, cultural narratives, knowledge, and the communication of ideas.

Conclusion

If, as is generally accepted, context is truly an important part of our interaction with, and reception and understanding of, cultural materials, it is remarkable that so little discussion concerning the entire range of contextual metadata to be recorded is found in the digital preservation literature. The original objects, whether digital or physical, are usually witnessed and/or exhibited in a way that offers some contextualization for our reuse and understanding of them. However, when physical objects are digitally preserved, they tend to be divorced from their original setting. De-contextualization is a fairly commonplace situation with cultural objects. A mechanism for capturing context that could be utilized within the preservation process would assist in the re-contextualization of the material for future use. Although gathering and preserving information to contextualize digital materials requires human effort, this work provides those interpretive narratives that are critical to successful use of materials in digital form. Because our world continues to embrace and depend on all things digital, ways to make sense of growing collections of preserved digital content is a difficult challenge that will need to be addressed. Without context the potential future usefulness of preserved digital content within the cultural heritage sector is limited.

Digitization permits individuals to interact with cultural objects in ways that were impossible just a few decades ago. While this is a boon to users, it must be remembered that the stories these objects tell are often impacted by differences between their physical and digital manifestations, and the passage of time between the digital content's creation and its later interpretation and reuse. Gaps in our knowledge of a cultural object's important attributes affect our understanding of its significance and its history. The work presented here identifies the various kinds of

information that bridge these contextual gaps.

Future work is planned to test the metadata framework[1]. It is hoped that through this work methods can be found to support the effective preservation of contextual information surrounding digital materials. If these efforts are successful, our understanding and reuse of these objects and our past will be greatly enriched.

Notes

[1] For the development of the framework and examples of its application see: Joan E. Beaudoin. (2012). A framework for contextual metadata used in the digital preservation of cultural objects. *D-Lib Magazine*, November 2012, 18(11/12). http://dx.doi.org/10.1045/november2012-beaudoin2

References

1. Adams, A. and Blandford, A. 2004. The unseen and unacceptable face of digital libraries. *International Journal of Digital Libraries*, 4, 71-81.
2. Bearman, D. 2007. Addressing selection and digital preservation as systemic problems. In Y. de Lusenet and V. Wintermans (Eds.) *Preserving the Digital Heritage: Principles and Policies*, 26-44. (Den Haag: European Commission for Preservation and Access).
3. Bénel, A., Egyed-Zsigmond, E., Prié, Y., Calabretto, S., Mille, A., Iacovella, A., and Pinon, J.-M. 2001. Truth in the digital library: from ontological to hermeneutical systems. *Lecture Notes in Computer Science*, 2163, 366-377. http://dx.doi.org/10.1007/3-540-44796-2_31
4. Benoit, G. 2011. Integrating use history as a context for dynamically updated metadata, *Journal of Library Metadata*, 11(3-4), 129-154. http://dx.doi.org/10.1080/19386389.2011.629958
5. Besek, J. 2003. *Copyright Issues Relevant to the Creation of a Digital Archive: A Preliminary Assessment*. (Washington, DC: Council on Library and Information Resources and the Library of Congress).
6. Besser, H. 2000. Digital longevity. In Maxine K. Sitts (Ed.) *Handbook for Digital Projects: A Management Tool for Preservation and Access*, 165-176.

7. Brocks, H., Kranstedt, A., Jäschke, G., and Hemmje, M. 2009. Modeling context for digital preservation. In E. Szczerbicki & N.T. Nguyen (Eds.) *Smart Information and Knowledge Management, Studies in Computational Intelligence*, 260, 197-226. http://dx.doi.org/10.1007/978-3-642-04584-4_9

8. Brown, J. S., and Duguid, P. 1996. The social life of documents. *First Monday*, 1(1).

9. Buckland, M. 1988. *Library Services in Theory and Context, 2nd ed.* (New York: Pergamon Press).

10. Bullock, A. 1999. *Preservation of digital information: issues and current status*. National Library of Canada, Network Notes #60.

11. Carrozzino, M., Scuccess, A., Leonardi, R., Evangelista, C., and Bergamasco, M. 2011. Virtually preserving the intangible heritage of artistic handicraft. *Journal of Cultural Heritage*, 12(1), 82-87. http://dx.doi.org/10.1016/j.culher.2010.10.002

12. Chen, S. 2001. The paradox of digital preservation. *Computer*, 34(3), 24-28. http://dx.doi.org/10.1109/2.910890

13. Chowdhury, G. 2010. From digital libraries to digital preservation research: The importance of users and context. *Journal of Documentation*, 66(2), 207-223. doi:10.1108/00220411011023625

14. Consultative Committee for Space Data Systems (CCSDS). 2009. *Reference Model for an Open Archival Information System (OAIS), Draft Recommended Standard*.

15. Conway, P. 2009. Building meaning in digitized photographs. *Journal of the Chicago Colloquium on Digital Humanities and Computer Science (JDHCS)*, 1(1), 1-18.

16. Conway, P. 1996. *Preservation in the Digital World*. (Washington, D.C.: Commission on Preservation and Access).

17. Copeland, A. and Barreau, D. 2011. Helping people to manage and share their digital information: a role for public libraries. *Library Trends*, 59(4), 637-649. http://dx.doi.org/10.1353/lib.2011.0016

18. Cowan, R.A., and Lillico, M. 2009. Increasing the value of university research records by preserving context. In: M. Pember & R.A. Cowan (Eds.) *iRMA Information and Records Management Annual 2009*. RMAA, St. Helens, Tasmania, 85-105.

19. Dalbello, M. 2004. Institutional shaping of cultural memory: digital library as environment for textual transmission. *Library Quarterly*, 74(3), 265-298. http://dx.doi.org/10.1086/422774

20. Day, M. 1997. Extending metadata for digital preservation. *Ariadne*, 9.

21. Duff, W., Monks-Leeson, E., and Galey, A. 2011. Contexts built and found: a pilot study on the process of archival meaning-making. *Archival Science*, 12(1), 69-92. http://dx.doi.org/10.1007/s10502-011-9145-2

22. Duranti, L. 2010. The long-term preservation of the digital heritage: a case study of universities institutional repositories. *Italian Journal of Library and Information Science*, 1(1), 157-168.

23. Duranti, L. (Ed.). 2005. *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*. (San Miniato, Italy: Achilab).

24. Duranti, L. 1995. Reliability and authenticity: the concepts and the implications. *Archivaria*, 39, 5-10.

25. Faniel, I.M. & Yakel, E. 2011. Significant properties as contextual metadata. *Journal of Library Metadata*, 11(3-4), 155-165. http://dx.doi.org/10.1080/19386389.2011.629959

26. Gilliand-Swetland, A. 2000. *Enduring Paradigm, New Opportunities: The Value of the Archival Perspective in the Digital Environment*. (Washington, DC: Council on Library and Information Resources and the Library of Congress).

27. Hedstrom, M. 1998. Digital preservation: a time bomb for digital libraries. *Computers and the Humanities*, 31,189-202. http://dx.doi.org/10.1023/A:1000676723815

28. Hirtle, P., Hudson, E., and Kenyon, A. 2009. *Copyright and Cultural Institutions: Guidelines for Digitization for U.S. Libraries, Archives, and Museums.* (Ithaca, NY: Cornell University Library).

29. Lagoze, C., Kraft, D., Payette, S., and Jesuroga, S. 2005. What is a digital library anymore, anyway? *D-Lib Magazine*. 11(11). http://dx.doi.org/10.1045/november2005-lagoze

30. Lavoie, B. and Gartner, R. 2005. *Technology Watch Report: Preservation Metadata*. OCLC; Oxford University Library Services.

31. Lavoie, B. and Dempsey, L. 2004. Thirteen ways of looking at ... digital preservation. *D-Lib Magazine*. 10(7/8). http://dx.doi.org/10.1045/july2004-lavoie

32. Lee, C. 2011. A framework for contextual information in digital collections. *Journal of Documentation*, 67(1), 95-143. http://dx.doi.org/10.1108/00220411111105470

33. Levy, D. M. 1998. Heroic measures: reflections on the possibility and

purpose of digital preservation. *Digital Libraries*, 152-161. http://dx.doi.org/10.1145/276675.276692

34. Lusenet, Y. de. 2002. *Preservation of digital heritage*. Draft discussion paper prepared for UNESCO.

35. Lynch, C. 2002. Digital collections, digital libraries and the digitization of cultural heritage information. *First Monday*, 7(5).

36. Marchionini, G. and Maurer, H. 1995. The roles of digital libraries in teaching and learning. *Communications of the Association of Computing Machinery*, 38(4), 67-75. doi: 10.1145/205323.205345

37. Mayer, R. & Rauber, A. 2009. Establishing context of digital objects' creation, content and usage. *InDP'09*, June 19, 2009, Austin, TX, USA.

38. McCarthy, G. 2007. Finding a future for digital cultural heritage resources using contextual information frameworks. In F. Camercon and S. Kenderdine, (Eds.) *Theorizing Digital Cultural Heritage: A Critical Discourse. (*Cambridge, MA: MIT Press), 245-260.

39. Meirelles, M. I.. (2004). Les CD-ROM Presence: The Ephemeral in Focus. *Proceedings of F@imp 2004 – International Audiovisual Festival on Museums and Heritage, Taipei, Taiwan*.

40. Menne-Haritz, A. and Brübach, N. n.d. *The Intrinsic Value of Archive and Library Material.*

41. Meyyappan, N., Al-Hawamdeh, S. and Foo, S. 2001. Digital work environment (DWE): using tasks to organize digital resources. *Lecture Notes in Computer Science*, 2163, 239-250. http://dx.doi.org/10.1007/3-540-44796-2_21

42. Nesmith, T. 2005. Reopening archives: bringing new contextualities into archival theory and practice. *Archivaria*, 60, 259-274.

43. Ormond-Parker, L. and Sloggett, R. 2011. Local archives and community collecting in the digital age. *Archival Science*, 12(2), 191-212. http://dx.doi.org/10.1007/s10502-011-9154-1

44. Palmer, C. 2002. Thematic research collections. Chapter in *Companion to Digital Humanities*.

45. Unsworth, J. 2004. *The value of digitization for libraries and humanities scholarship*. Innodata Isogen Symposium.

46. Vogt-O'Connor, D. 2000. Selection of materials for scanning. In Maxine K. Sitts (Ed.) *Handbook for Digital Projects: A Management Tool for Preservation and Access*, 45-72.

47. Waibel, G. 2003. Like Russian dolls: nesting standards for digital preservation. *RLG DigiNews*, 7(3).

48. Wallis, J.C., Borgman, C.L., Mayernik, M.S., & Pepe, A. 2008. Moving archival practices upstream: an exploration of the life cycle of ecological sensing data in collaborative field research. *International Journal of Digital Curation*, 1(3), 114-126. http://dx.doi.org/10.2218/ijdc.v3i1.46

49. Watry, P. 2007. Digital preservation theory and application: transcontinental persistent archives testbed activity. *International Journal of Digital Curation*, 2(2), 41-68. http://dx.doi.org/10.2218/ijdc.v2i2.28

50. Wisser, K.M. 2011. Describing entities and identities: the development and structure of encoded archival context—corporate bodies, persons, and families. *Journal of Library Metadata*, 11(3-4), 166-175. doi: 10.1080/19386389.2011.629960

About the Author

Joan Beaudoin is an Assistant Professor in the School of Library and Information Science at Wayne State University where she teaches and performs research on metadata, information organization, digital libraries, digital preservation and visual information. Prior to her position at Wayne State University she was a Laura Bush 21st Century Librarian Fellow at the School of Information Science and Technology at Drexel University. In addition to a Doctor of Philosophy in Information Studies at Drexel University, she holds a Master of Science in Library and Information Science degree in the Management of Digital Information from Drexel University, a Master of Arts in art history from Temple University, and a Bachelor of Fine Arts in art history from Massachusetts College of Art.