The current issue and full text archive of this journal is available at
www.emeraldinsight.com/0022-0418.htm

A framework for contextual information in digital collections

Christopher A. (Cal) Lee
University of North Carolina, Chapel Hill, North Carolina, USA

Received 1 September 2009
Revised 13 April 2010
Accepted 19 April 2010

Abstract

Purpose – This paper sets out to investigate the meaning, role and implications of contextual information associated with digital collections.

Design/methodology/approach – This paper is based on an extensive review and analysis of both the scholarly literature from many disciplines about the concept of context and the professional literature (including standards) related to the description of information artifacts. The paper provides an analysis of context, distinguishing three main ways in which that term has been used within the scholarly literature. It then discusses contextual information within digital collections, and presents a framework for contextual information. It goes on to discuss existing standards and guidance documents for encoding information related to the nine classes of contextual entities, concluding with a discussion of potential implications for descriptive practices through the lifecycle of digital objects.

Findings – The paper presents a framework for contextual information that is based on nine classes of contextual entities: object, agent, occurrence, purpose, time, place, form of expression, concept/abstraction, and relationship.

Research limitations/implications – Research and development about and in support of digital collections will benefit from a clear articulation of the types, roles, importance and elements of contextual information.

Practical implications – Future users of digital objects will probably have numerous tools for discovering preserved digital objects relevant to their interests, but making meaningful use and sense of the digital objects will also require capture, collection and management of contextual information.

Originality/value – This paper synthesizes and extends a previously diffuse literature, in order to clarify and articulate core concepts in the management of digital collections.

Keywords Collections management, Metadata

Paper type Research paper

> Always design a thing by considering it in its next larger context – a chair in a room, a room in a house, a house in an environment, an environment in a city plan – Eliel Saarinen (The Maturing Modern, 1956).

> . . . if life is going to exist in a Universe of this size, then the one thing it cannot afford to have is a sense of proportion – Douglas Adams (1980).

1. Introduction

Numerous forms of expression and social interaction take place through digital media. Having access to traces of these expressions and interactions will be essential for future users to know about, appreciate and understand the details of our current lives.

This work has been supported, in part, through NSF Grant No. IIS 0455970. Jane Greenberg, Jennifer Engleson Lee, Laura Sheble, Paul Solomon and Helen Tibbo have all provided valuable suggestions.

Journal of Documentation
Vol. 67 No. 1, 2011
pp. 95-143

0022-0418
DOI 10.1108/00220411111105470

Future users will have numerous tools for discovering preserved digital objects that are relevant to their particular interests, but this does not mean that they will be able to make sense of the digital objects.

Literature about the curation of digital collections frequently cites the importance of reflecting context associated with digital objects. However, there has been relatively little detailed discussion of:

- what "contextual information" means; and
- how curators of digital collections might best create, capture, encode, manage and provide access to contextual information.

This paper addresses both questions[1].

2. What is context?

Broadly speaking, context is "the circumstances that form the setting for an event, statement, or idea" (*Compact Oxford English Dictionary of Current English*, 2005). Context is inherently relational; it is always context of, about, or surrounding something, which I will call the target entity (TE). In relation to a given TE, the broadest formulation of context would be "everything else," i.e. everything (states, objects, facts, relationships) in the universe that is not the TE. If one wanted to know the full context of an entity, one would need an omniscient awareness of all existence. Such a conception of context, however, would not be very useful, nor would it reflect the "thrownness" (Heidegger, 1996; Winograd and Flores, 1986), "embodiment" (Lakoff and Johnson, 1999; Dourish, 2001), or "situatedness" (Lindblom and Ziemke, 2003) of the human condition – acting, thinking, learning, developing, caring and perceiving within, about and through a particular lived subset of existence – or the insights from psychology about figure-ground relationships, which imply that the meaning of a given state can depend dramatically on one's particular focus of attention. Contexts can "act like adjustable filters for giving the right meaning" (Brézillon, 1999, p. 49). A more human-centered version of the "everything else" definition could be "anything that is not defined as the phenomenon of interest" (Dervin, 1997). Glaser and Strauss (1964) offer a slightly more constrained notion of context as "a structural unit of an encompassing order" that is "larger than" and "surrounds and affects" the "unit under focus." Brézillon calls it "what constrains problem solving without intervening in it explicitly" (Brézillon, 1999, p. 48).

There are no absolute rules for determining *a priori* what will count as the context of a TE (Greenberg, 2001), but something is generally more likely to be considered part of the context if it is "proximate" (Guha and Lenat, 1994) to the TE along a particular dimension or for a particular purpose. In short, context is a set of things, factors or attributes that are related to a TE in important ways (e.g. operationally, semantically, conceptually, pragmatically) but are not so closely related to the TE that they are considered to be exclusively part of the TE itself.

Within a particular conversation, discipline or school of thought, the boundary between (i.e. what should be considered part of) the following three categories is a matter of ongoing negotiation and evolution:

(1) TE;

(2) context of the TE; and

(3) things not relevant enough to be considered part of either (1) or (2).

Within a given area of research, "context is some kind of a background for something the researcher wishes to understand and explain" (Talja *et al.*, 1999). The boundary between content and context is "pragmatic, permeable and revisable" (Callon and Law, 1989) and "is continually negotiated and re-negotiated" (Lea *et al.*, 1995). Within research communities, these distinctions are often closely connected with decisions about units of analysis. Many social scientists, for example, have emphasized the need to attend to surrounding elements, in order to understand the TE, through units of analysis such as activity (Kaptelinin and Nardi, 2006), practice (Bourdieu, 1977), domain (Hjørland and Albrechtsen, 1995), situated action (Suchman, 1987), scene (Fillmore, 1977; Blum, 2003; Tyler and Evans, 2003), situation (Dervin, 1983), episode (van Dijk, 1981), and setting (Goffman, 1959; O'Donnell *et al.*, 1993).

Across a variety of disciplines, specific formulations of context tend to emphasize one (or more) of the following:

- *Context*$_1$ – the set of symbolic expressions or representations that surround a TE and help one to express, make sense of, translate or otherwise act upon or within it.
- *Context*$_2$ – objective or socially constructed characteristics and conditions of the situation in which a TE is, appears or occurs.
- *Context*$_3$ – aspects of the mental or physical state, disposition, intentions, identity or recent experiences of an actor that bear upon how she interprets, understands, acts within, or what she notices of, the situation at hand.

The first meaning of context (context$_1$) is about a TE's place within a larger discourse or information system. Examples include the discourse within which a statement is embedded; other documents filed in the same category; and a formal theory within which a concept or statement is to be understood. The conveyance of meaning between a TE and its context$_1$ is often bi-directional; if the information from a TE is asserted or conveyed to those involved in a given context$_1$, "then that information will become part of the body of information that provides the context for the subsequent discourse" (Stalnaker, 1998, p. 5). Contextual analysis – the analysis of surrounding text of a work in order to make sense of it – is an example of an activity that is based on context$_1$.

The second (context$_2$) is about the objective or inter-subjectively recognized set of factors surrounding a TE[2]. Examples include location; temperature; being under water; occurring as part of a traditional ritual[3]; position within the reporting structure of an organizational hierarchy; relative arrangement and orientation of objects; existence and accessibility of other surrounding objects[4].

The third (context$_3$) is about the subjective status of a particular agent – e.g. a user in "context-aware computing" (Schilit and Theimer, 1994), one ascribing knowledge to a statement in epistemology, a participant in a speech act. This type of context includes not only "where I stand" and "what I'm currently thinking" but also what has been variously called fringe (James, 1890), horizon (Husserl, 1952; Heidegger, 1996; Gadamer, 1989), habitus (Bourdieu, 1977), or habit (Dewey, 1922) – all of which emphasize and attempt to explain the intimate connections between agency and the world in which agents are embedded. In their discussion of multi-sensory communication, Mani and Sundaram (2007) break down context$_3$ further as either context of construction or context of interpretation (context for the message transmitter or receiver, respectively). "Common ground" is shared context$_3$ that allows two or more individuals to understand each other (Stalnaker, 1978; Clark, 1996). The confluence of

contexts[3] (e.g. through conversations or other interactions between individuals; or interaction between an individual and a document) can be characterized as a "fusion of horizons" (Gadamer, 1989). From the perspective of a given agent (A), $context_3$ can take the form of:

- A's own state, disposition or
- the state, disposition of other agents that are relevant to the matter at hand[5].

Borrowing terminology from sociology and economics, we could call:

- "ego $context_3$"; and
- "alter $context_3$."

Glaser and Strauss's (1964) concept of "awareness context" is a subset of $context_3$: "the total combination of what specific people, groups, organizations, communities or nations know what about a specific issue." Two similar concepts are the "frame," which is "a remembered framework to be adapted to fit reality by changing details as necessary" (Minsky, 1974) and the "primary framework," which is a way in which one's attention is channeled in order to render "what would otherwise be a meaningless aspect of the scene into something that is meaningful" (Goffman, 1974). Frames and the process of framing – placement within one's interpretive context – have played a significant role in studies of politics, social movements, communication and media (Snow *et al.*, 1986; Scheufele, 1999; Lakoff, 2008)[6]. Because it involves an intentional agent, $context_3$ reflects not only aspects of the past, but also aspects of the future (as expected, hoped, predicted or planned by the agent). This is reflected in the notion of "horizon" as a coming together of past, present and future.

Closely related to all three types of context is the importance of history and "path dependence" (Liebowitz and Margolis, 2000); the relevant context often includes aspects of previous states of things. As with context in general, there are no absolute rules for whether something from the past counts as the context of a contemporary TE – what Dervin (1997) calls the "width of history" – but there is a rough correlation between something being part of the context and its relative recency (proximity in time to the TE). For example, if Alice and Betty are engaged in a conversation, a statement that Alice made five seconds ago is more likely to serve as $context_1$ for Betty's current statement than is a statement that Alice made five hours ago. In a game of billiards, the $context_2$ for my next shot is more clearly defined by my opponent's previous shot than it is by a shot someone took yesterday on the same table. Likewise, the $context_3$ in which one interprets a new experience is more likely to be defined by her recent thoughts and experiences than by an experience from many years ago.

A great deal of human communication takes place at the intersection between the three types of context. In a spoken or written declarative statement (T), a person (S) attempts to convey to other individuals some objective aspects of the world ($context_2$) or something that she has personally experienced or is currently experiencing (her $context_3$). Those who encounter T can then makes sense of T by using a combination of the words that constitute T, other related information entities that are at hand ($context_1$), characteristics of the current state of things ($context_2$), and their own repertoire of knowledge or experiences at the time ($context_3$). Weick (1995) argues that context or "local contingencies" influence not only how things are interpreted but also what is noticed or "extracted as a cue in the first place."

Several authors provide taxonomies of context, which get at the dynamic interchange between artifacts, symbols and a surrounding set of conditions or activities. For example, Léger *et al.* (2005) distinguish task, visual and semantic context; and Ingwersen and Järvelin (2005) identify intra-object, inter-object, social, organizational, cultural and systemic context. Agre (2001) distinguishes between architectural and institutional aspects of context, and he discusses challenges that emerge when the two are not clearly aligned. The InterPARES project identifies juridical, administrative, provenancial, procedural, documentary, and technological context (*Long-Term Preservation of Authentic Electronic Records*, 2002).

Formal information systems (e.g. theories,[7] databases, library catalogs) also act at the intersection of $context_1$, $context_2$ and $context_3$. Many of the conventions for activities that build and maintain our "epistemic infrastructure" (Hedstrom and King, 2006) – e.g. scholarly research, developing inventories of collections – are attempts to surround expressions about aspects of the world ($context_2$) with sufficient and appropriate supporting information ($context_1$) so that they will be interpreted by future readers or listeners in (at least relatively) predictable ways.

Reflecting previous contexts is often considered a matter of professional responsibility. A "point of sharp contrast between the archaeologist and the looter," for example, is that the latter "does not bother to record" contextual information before removing an object from the place it was found (Sharer and Ashmore, 1987). An ideal set of descriptive information ($context_1$) associated with an archeological artifact would allow a user of that associated information to mentally reconstruct all relevant aspects of the environment in which the artifact was found ($context_2$). An essential consideration is which aspects one considers to be "relevant." No set of information objects, no matter how detailed, can fully capture and reflect everything about a given situation ($context_2$). Instead, archeologists establish and perpetuate principles and heuristics that guide their conveyance of context.

Issues of interoperability (e.g. hardware and software required to store and render a particular file) are outside the scope of this paper. However, there are aspects of technical/technological context (*Long-Term Preservation of Authentic Electronic Records*, 2002; Suderman, 2001) that a future user may need to know in order to adequately understand and use a given digital object. This could include "requirements, capabilities, limitations, design, operation and maintenance of the creator's original system" (Canadian Council of Archives, 2003). Information about the availability, use and functionality of particular applications or computing platforms in a particular time and place "will become even more important as users, over the course of time, become less and less likely to have had firsthand knowledge or experience with obsolete computing platforms" (Hedstrom *et al.*, 2006).

3. Contextual information in digital collections

It is important to distinguish between the broad notion of context – constituted by the interactions and relationships between a TE and its environment – and the more specific set of contextual information that is reflected in information systems. There will always be limits to what any representation system – whether that system is in someone's head, embedded in a digital object or within the descriptive apparatus surrounding a digital object – can reflect about the environment in which it was originally embedded (Shanon, 1990). Those responsible for designing, implementing

and managing information systems are in the business of using symbolic representations or collections of symbolic representations (a form of $context_1$) in order to capture and maintain relevant aspects of $context_2$ and $context_3$. "Context, in principle, is infinite. The describer selects certain layers for inclusion, and decides which of those to foreground" (Duff and Harris, 2002). As Heraclitus tells us, it is impossible to step into the same river twice. No two situations ($context_2$) will be exactly the same, though two situations will often be identical for all practical purposes. Humans tend to operate under the "default [assumption] that most contextual factors don't affect the representation of most facts. This is the default that allows us to assume that most of what we know applies even in completely new and strange situations" (Guha and McCarthy, 2003). An essential condition for communication between humans is a shared "commonsense knowledge" that allows most aspects of context to go unnoticed and unmentioned. When interacting directly with each other in a shared time and place, humans "make handling context look easy. This can lead us to underestimate the subtleties and complexities of digitally acquiring, representing, and acting on contextual information" (Grudin, 2001). The "problem of how general to be [when specifying context] arises whether the general commonsense knowledge is expressed in logic, in program, or in some other formalism" (McCarthy, 1987).

Likewise, no digital object can carry all of its context along with itself. In Leibniz's terms, there is no such thing as a digital monad, i.e. a fully self-contained, self-describing digital object that represents the entire universe (full elaboration of all three types of context) that surrounds itself. Lifting a digital object out of its original context in order to be used in another context carries with it dangers of both omission and commission: without access to sufficient contextual information, a user can suffer from gaps in understanding, but also based on the natural human propensity to make sense and reduce cognitive dissonance (Weick, 1995), the user is likely to mentally "fill in the gaps" based on characteristics of her current context. According to Dewey (1931), "The greater the degree of remoteness, the greater is the danger that a temporary and legitimate failure to express reference to context will be converted into a virtual denial of its place and import."

Contextual information can help to reflect "the organizational, functional, and operational circumstances surrounding materials' creation, receipt, storage, or use, and its [*sic*] relationship to other materials" (Pearce-Moses, 2005). Documents can derive considerable value and meaning from their relationships with other documents within the same collection. "In order to understand any object and its significance, the person experiencing it must have a context to set it in" (Allison *et al.*, 2004).

Relationships to other digital objects can dramatically affect the ways in which digital objects have been perceived and experienced. In order for a future user to make sense of a digital object, it could be useful for that user to know precisely what set of surrogate representations – e.g. titles, tags, captions, annotations, image thumbnails, video keyframes – were associated with a digital object at a given point in time. It can also be important for a future user to know the constraints and requirements for creation of such surrogates within a given system (e.g. whether tagging was required, allowed, or unsupported; how thumbnails and keyframes were generated), in order to understand the expression, use and perception of an object at a given point in time.

Placement and arrangement of digital objects in relation to each other can also serve as an important form of contextual information (Hedstrom and Lee, 2002). In the archiving of web pages, for example, if the portions of a page's frameset are crawled at

different times, a future user could get a false impression of what really appeared on the page at a given point in time (Foot *et al.*, 2003). Ranking and classification of individual objects can also influence the ways in which they are presented to and experienced by users. In many online environments, for example, contributors have been accused of creating false subscriptions for the sole purpose of raising their number of views, in-links and subscribers, or using software to "view" a page many times, thus raising their visibility within the system. Understanding these factors could be essential for future users to understand how a given object was actually presented, used and perceived.

4. Building-blocks and motivation for the contextual information framework

In developing the framework in this paper, I have drawn from a diverse range of sources. The most direct guidance has come from high-level ontologies; descriptive and metadata standards, which specify classes, entities, elements, properties and attributes that can serve as contextual information for a target object; and standards that define and potentially support the interactions between a diverse set of entities over time. See Appendix for an elaboration of sources consulted and their relationships to different types of contextual entities.

Since the nineteenth century, the literature on cataloging and classification has reflected systematic efforts to identify and disentangle the various entities related to library materials. Given practical concerns about where to place items on shelves and how to find items related to given research questions, the cataloging literature has traditionally placed significant emphasis on what specific entity or entities to include in the main or added catalog entries for a item, and how to structure classification systems so their serialized representations (e.g. call numbers on the sides of the books) will support search, discovery, browsing and addition/removal of categories over time. Major access points for a book have tended to be:

- who it is by;
- when it was published; and
- what it is about[8].

The cataloging literature provides the most depth of thought and detail regarding how to make assertions of an item's aboutness[9]. Early classification systems tended to be enumerative, i.e. articulating a full list of all allowable subjects. During the twentieth century, there were numerous efforts to develop faceted classification systems that do not assign "fixed slots to subjects in sequence" but instead use "clearly defined, mutually exclusive, and collectively exhaustive aspects, properties, or characteristics of a class or specific subject" (Taylor and Miller, 2006, p. 394). Beginning in the 1930s, S.R. Ranganathan provided a very detailed articulation and rationale (Ranganathan, 1938) for faceted classification, which is reflected in the Colon Classification system (Ranganathan, 1965). Ranganathan identified five fundamental categories: time, space, energy, matter, and personality (Ranganathan and Gopinath, 1967). Efforts of the Classification Research Group, as well as others involved in the development of the Second Edition of the Bliss Bibliographic Classification, have resulted in a more detailed elaboration of 13 categories to serve as the basis for facets: thing, kind, part, material, property, process, operation, product, by-product, patient, agent, space, and time (Broughton, 2010).

Regardless of how capable a given bibliographic classification system is to evolve and accommodate changes in systems of knowledge over time, its application (formal assertion of an item's aboutness) generally takes place at a particular point in time when someone creates a catalog record for the item. The contextual information framework that I present in this paper is intended also to fulfill two further needs: documenting the traces of entities that have numerous other relationships (beyond aboutness) to a target digital object and documenting the states and characteristics of the entities themselves over time. These two needs complement, but are distinct from, making specific assertions about the entities and their relationships for a particular item at a given point in time.

Another source of guidance is the literature related to conveying "circumstances" for rhetorical purposes. In the first century BCE, Hermagoras articulated circumstances as "*quis, quid, quando, ubi, cur, quem ad modum, quibus adminiculis*" (Robertson, 1946), which can be roughly translated as who, what, when, where, why, in what manner, and by what means. Over the past two millennia, there have been various articulations and adaptations of similar lists of questions, including those widely taught to journalists and investigators: who, what, where, when, why and how.

The professional literature of archivists provides many valuable components. When classifying public records, Schellenberg (1956) suggests that the three main elements to consider are "the action to which the records relate, the organizational structure of the agency that produced them, and their subject matter" (Schellenberg, 1956, p. 53), and he further indicates the analysis of government transactions reveals relationships to persons, corporate bodies, places, and topics (p. 54). Bearman (1989) proposes seven "fundamental dimensions": space, time, subject, action, object, form and function. He further suggests, "User queries are intended to retrieve those representations of realities which have something in common along one or more selected dimensions" (Bearman, 1989, p. 50).

Literature about information needs in various user populations suggests the value of identifying contextual elements related to the objects in a collection. The particularly important role of proper names (people and places) in the research of humanities scholars is reported in Wiberley and Jones (1989) and Buchanan *et al.* (2005), who also report the use of chronological periods. In her study of abstracts in history, Tibbo (1993) identifies the importance of time, place, historical players and events. Bates *et al.* (1993) (Bates, 1996) found that humanities scholars emphasized named individuals, geographical terms, chronological terms, and discipline terms (as opposed to more specific subject headings). Cole (2000) found that the discovery, identification and collection of information related to particular names of people, places, and things played an important role in the research of history doctoral students. In their analysis of email reference questions, Duff and Johnson (2001) identify the following elements: proper names, dates, places, subject, form, and, events. In queries to ABC-Clio, Yi *et al.* (2006) found instances of queries based on historical events, people, and regions. Chin and Lansing (2004) identify the most prominent "scientific and social contexts in which data is created, interpreted, and applied" among biologists to be: general data set properties; experimental properties; data provenance; integration; analysis and interpretation; physical organization; project organization; scientific organization; task; experimental process; and user community. Yang and Marchionini (2005) identify nine "visual gist attributes" for making sense of a video after viewing a fast forward

surrogate: action/activities/events, geographical location, object, people, plot, setting/environment, theme/topic, time/period and visual perception. Henderson identifies four facets used by individuals to name folders for managing their documents: genre, task, topic and time (Henderson, 2005). Contextual information is useful not only for navigation and discovery of relevant items but also for understanding and making sense of items once they have been found. For example, Sweet and Thomas (2000) explain that collection-level description allows users who have found individual documents to "then move 'bottom upwards' to see the context in which the documents were created and used."

The Reference Model for an Open Archival Information System (OAIS) (ISO 14721, 2003) defines context information as "the information that documents the relationships of the Content Information to its environment. This includes why the Content Information was created and how it relates to other Content Information objects." According to the OAIS, digital objects should be managed in such a way that they will be "independently understandable" to a "Designated Community." The farther away in time, place, and social situation a Designated Community is from those who originally created and used a digital object (or set of digital objects), the less likely it is that the following will be available to the Designated Community in order to use and make sense of the digital object(s): associated documentary material ($context_1$), similar characteristics of the world ($context_2$), and commonality with the resources' creators, in background experience, perspective, or knowledge ($context_3$).

Contextual information is often characterized by recursive relationships. A component of information that was created or captured to provide contextual information about a TDO can often itself become a TDO, about which one may create, capture, and manage additional contextual information (see Figure 1). Like most recursive relationships, there is always a practical limit to the recursion. Holdsworth

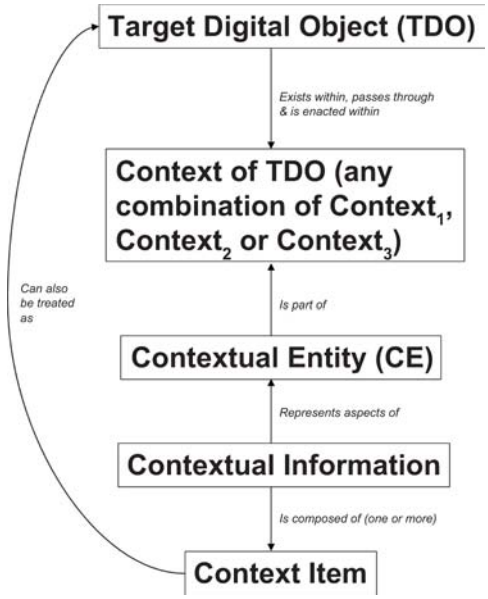


Figure 1.
Relationships between target digital object, context, contextual information, and context item

and Wheatley (2004) use the label "Gödel ends" for references to information outside the control of a repository. They explain that such references are "an inevitable consequence of Gödel's incompleteness theorem" and often "relate to the current practice of the time." The final arbiters of contexts are human agents. In the sharing of scientific data, for example, there are definite limits to how many layers of contextual information one can specify in an information system. The majority of rich contextual information about data takes the form of tacit knowledge of researchers (Zimmerman, 2008) and is often conveyed through individuals and communities of practice (Birnholtz and Bietz, 2003) rather than formal information systems.

5. Proposed contextual information framework

Recall that context is always the context of some target entity (TE). The framework presented in this paper is intended to inform the creation, capture and curation of the contextual information within a repository, which can help to understand, make sense of, analyze and use a particular target digital object (TDO). Note that this is different from a conception of context that places the user or particular use in the center of the picture, as one does when talking about "context-sensitive help," which changes the interface based on the current state of an application or "context-aware computing" (Schilit and Theimer, 1994), which attends to the current location and configuration of a mobile device.

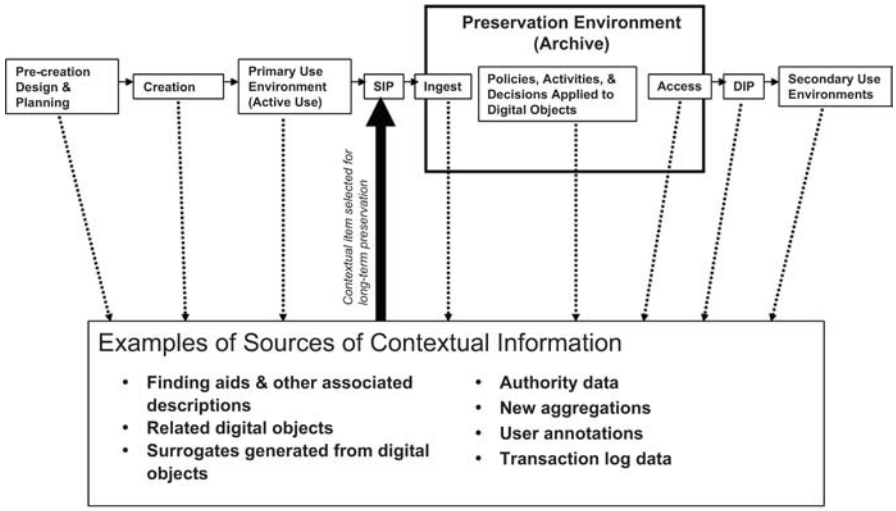
5.1 Basic definitions

The TDO is the digital object that is being explained or understood through contextual information. A contextual entity (CE) is something in the world that could be related to a TDO as part of its context. The main criterion for whether something counts as a CE associated with a given TDO can be stated as a conditional statement: If a user of the TDO were exposed to more information about the CE than the digital object itself provides, the user would better understand something about the TDO's context. Building on the terminology of the Java Context Awareness Model (Bardram, 2005), I use the phrase "context item" to refer to a digital object (e.g. string of text, image, video segment, file, record in a database) within a repository that carries information about a CE. If there is contextual information associated with a TDO, it will be composed of one or more context items. A given context item can provide contextual information about one or more CEs.

Throughout its existence, a TDO could pass through innumerable contexts[2], most of which will not be noticed by anyone nor will they be explicitly documented in an information system. Part of the work of a digital curator is to determine what points in the life of digital objects (i.e. which contexts in which it has been) should be explicitly identified and described within the archive. In order to ensure that information about a given CE (or context through which a digital object passes) is to persist over the long term, the information should be embedded in a context item that is ingested into a trustworthy preservation environment (see Figure 2).

5.2 Classes of contextual entities

There are innumerable ways in which one can classify aspects of the universe, each of which could be particularly appropriate to a given purpose. Section 4 discussed sources that have served as building blocks for this framework by providing entities that were created for other purposes. In developing the contextual information framework presented in this paper, my purpose has been to elaborate the minimum number of



Notes: SIP = Submission Information Package (OAIS); DIP = Dissemination Information Package (OAIS). Dashed arrows = generation of contextual information. Solid arrow = selection of contextual information for long-term preservation within preservation environment (Archive)

Figure 2.
Potential generation and preservation of contextual information in the life of a digital object

categories of contextual entities that would be required to comprehensively[10] document the "life history" of a target digital object.

Each class within the framework identifies a type of contextual entities that can have their own set of "persistent state information,"[11] which can:

- help a user agent to better understand the context of a target digital object;
- "be distinctly identified" (Chen, 1976);
- be expressed using a fixed representation (physical artifact or binary-encoded symbols); and
- not be fully expressed through the persistent state information of other types of entities.

Table I presents the nine classes of contextual entities that are intended to meet the above conditions.

One way of understanding the different roles played by the nine types of entities is through analogy to the parts of a declarative sentence. An object is a noun that lacks intentionality (e.g. book, file, tree, house). An agent is a noun that has intentionality either individually (person) or as a collective unit (e.g. organization, family, club, nation, corporation). An occurrence is something happening (e.g. election, rocket launch, hurricane, war, collection of a data element as part of a social science survey, conversion of a file from one format to another). A purpose is a reason or motivation for an actor to engage in an event (e.g. legal mandate, teaching objective, organizational program area). Time and place provide the spatio-temporal boundaries within which objects and agents reside and events occur. A concept or term is an abstraction or idea. Form of expression characterizes the way the sentence, or some part of it, is expressed.

Object	A bounded discrete entity that can be characterized as having one or more properties or states; persist across multiple points in time and place; be uniquely identified; interact with other objects; and be acted upon by an agent
Agent	An entity that can carry out actions
Occurrence	A characterization, for a given span of times and places, of either the state of a set of entities or their interaction(s)
Purpose	Mandate, norms, values, intention, rules, standards, virtues, or functions to which agents can advance or with which they can conform; attempt to advance or conform; hope to advance or conform; or perceive/expect entities (or sets of entities) to advance or conform
Time	"A limited stretch or space of continued existence, as the interval between two successive events or acts, or the period through which an action, condition, or state continues" (*Oxford English Dictionary*, 1989)
Place	A designated point or region in space
Form of expression	A particular way of expressing ideas or information
Concept or Abstraction	Ideas or other individually/socially recognized "properties or qualities as distinguished from any particular embodiment of the properties/qualities in a physical medium" (*Standard Upper Merged Ontology*)
Relationship	An association between two or more entities (or classes of entities), which cannot be reduced to or adequately expressed as a property of the entities (or classes of entities) themselves

Table I.
Nine classes of contextual entities

Consider the following case: As part of its K-12 education initiative [purpose], NASA [agent] hired [occurrence] ABC Corporation [agent] as a contractor [relationship] in 2003 [time] in Greenbelt, Maryland [place] to create the movie, "Sonic Boom" [target digital object] and an associated educational web page [object], which explain sonic booms [concept].

A given context item can provide contextual information about one or more CEs. Lagoze *et al.* (2005) describe this as "polymorphism" whereby "a digital object may assume any combination of type identities." There is no inherent property of a digital object that determines whether or not it provides contextual information about another object, agent, etc. It is not always possible, based simply on its name, to infer what class of entity is being contextualized by a given CI. Given the use of metonymy in natural language, for example, "September 11" could stand for either September 11, 2001 or a set of events that occurred on that day in the USA (Ferro *et al.*, 2005), and "the White House" could be an object, agent, place, or purpose.

5.2.1 Object. An object is a bounded entity that can:

- be characterized as having one or more properties or states;
- persist across multiple points in time and place;
- be uniquely identified;
- interact with other objects; and
- be acted upon by an agent.

According to Smith (1996), "an individual object is taken to be something of coherent unity, separated out from a background, in the familiar 'figure-ground' fashion." For the purpose of digital curation, the target object is the thing being preserved and managed, which is a "content bearing object" (Niles and Pease, 2001). Objects can be atomic units or they can have internal complexity. Three important properties of a digital object are version (e.g. draft vs final; master vs service copy), level of abstraction (e.g. logical object vs. particular instance)[12], and level of aggregation (e.g. repository, collection, sub-collection, series, information package, digital object, object component/segment).

A collection is a type of object, which is "a grouping of individual items or other collections" (Brack *et al.*, 2000). Collections are usually defined by relationships (see section 5.2.9 below) between one or both of the following:

- the objects that make up the collection (e.g. similarity in topic or documentary form); or
- the objects and some other entity or category of entities (e.g. common provenance, intended audience).

An object can be part of more than one collection. For any given complex digital object, one could decide to treat the object as a collection of its components (with those components then being treated as objects). Conversely, anything considered to be a collection could instead be treated as a single object (e.g. an entire data set being managed as one atomic unit). In practice, curators in a particular situation will often have established conventions for what level they consider to be a collection and what level they consider to be an object within a collection. One advantage of treating a collection as an entity is that it can be managed separately from the objects it contains. This can then reflect that a given object may have been treated as part of a given collection in the past but is now considered part of a different collection.

5.2.2 Agent. Agents are entities that can carry out actions. This includes not just individuals but also organizations, divisions and programs within organizations. Agents are different from objects in that they have:

- some degree of intentionality (see purpose below); and
- the ability to use that intentionality to cause changes in the state of other entities.

In short, an agent is one who can purposely do things. The Dublin Core Metadata Initiative (DCMI) Agents Working Group has defined agents more narrowly as either persons or groups that "have a role in the lifecycle of a resource" (Wilson and Clayphan, 2004). When documenting the custodial history of a target digital object, it can also be important to identify agents (e.g. "archival institutions and repositories") who were previous responsible for the object (Vitali, 2006).

The attributes of agents can vary dramatically in the degree to which they persist over time. As a result, the capture and preservation of some attributes will be much more time-sensitive than others (implying the need to capture them before they are lost). For example, one's social security number is much less likely to change from one year to the next than is one's weight, salary or occupation. On even shorter time scales, "emotional states hold normally no longer than 15 minutes, however personality traits won't change within months" (Heckmann *et al.*, 2005).

The boundary between objects and agents is one of convention[13]. One may decide to attribute agency to only certain categories of living organisms (e.g. only humans, mammals, vertebrates, or animals). One may also decide to attribute or fail to attribute agency to software or machinery that carries out actions on behalf of individuals or groups. PREMIS, for example, defines an agent as "a person, organization, or software program associated with preservation events in the life of an object." The MPEG Rights Expression Language uses the term "principal" which can be a person, group, software application, device or other entity potentially authorized to use resources. Computer science and artificial intelligence literature often uses the term "agent" to refer to a piece of software or cluster of processes that someone can intentionally enroll to act on one's behalf, even if the agent does not itself have intentionality, e.g. Minsky's (1986) conception of the mind as a society of agents, or "user agents" that request and retrieve information over the internet. In many situations, it can be important to distinguish between conscious actors (human beings) and the devices that do things for them. The current draft of functional requirements from DCMI Agent Working Group, for example, no longer includes the category of "automaton (weather recording device, software translation program, etc.)" which was part of the group's earlier definitions. Kaptelinin and Nardi (2006) identify six different "forms of agency":

(1) natural things;
(2) cultural things;
(3) natural nonhuman living beings;
(4) cultural nonhuman living beings;
(5) human beings; and
(6) social entities.

The boundary between agents and concepts is also one of convention. Various expressions of personal or group identity, for example, could be treated as either agents, attributes of other agents, or concepts. These include personae, pseudonyms, aliases, pen names, ring names, profiles and avatars, all of which project some aspects of an agent's (or multiple agents, in the case of shared identity) personality but often in a limited or deceptive manner. Fictional characters and supernatural entities can also be considered either agents or concepts.

5.2.3 Occurrence. An occurrence is a characterization, for a given span of times and places, of either the state of a set of entities or their interaction(s). Occurrences are usually transformative, i.e. the entities involved in the occurrence are somehow different after it has occurred. Types of occurrences are events, processes, actions, activities, and accomplishments (Allen, 1984). Stated more simply, occurrences are "situations that *happen* or *occur*" (Pustejovsky *et al.*, 2003).

A process is "a naturally occurring or designed sequence of changes of properties or attributes of an object or system" (Wikipedia, 2009) – or "a series of actions or events taking place in a defined manner leading to the accomplishment of an expected result" (ISO/IEC 15944-1, 2002) – whereas an event is a more specific "transition from one situation to another" (Lagoze and Hunter, 2001) in which "the state of the world changes" (Matuszek *et al.*, 2006). Activity within an organization is often associated with a business process, which can be defined as "a collection of interrelated activities,

initiated in response to a triggering event, which achieves a specific, discrete result for the customer and other stakeholders of the process" (Sharp and McDermott, 2009).

Events tend to be "countable" (Mourelatos, 1978), whereas processes can reach across indefinite spans of space and time, often making them difficult to count. Doerr and Kritsotaki characterize events in terms of "meetings," which are "interactions of living or dead items that bring about changes of state" (Doerr and Kritsotaki, 2006). "Landmarks" are particularly noteworthy events that can trigger memories, guide discovery and facilitate personal information management tasks (Ringel *et al.*, 2003). Dewey says, "every occurrence is a *con*currence" of elements in an ongoing flow of existence, i.e. "it is inherently characterized by something from which and to which" (Dewey, 1931), so any characterization of something as a completely discrete event is a simplification for purposes of representation.

Whether one considers a given occurrence to be a process or an event will depend on the level of abstraction that seems most appropriate. For most purposes, for example, one would consider a flash of lightning to be an event, but less discrete occurrences involving groups of agents (e.g. the Civil War) or individuals (e.g. "coming of age") would be considered either processes or "long-running events" (Smith, 2002).

Occurrences (both processes and events) can take the form of either general phenomena in which there is no specific acting entity, e.g. Hurricane Katrina (unless one wishes to identify God or Mother Nature as an agent), or actions, which are events carried out by some identifiable agent (or set of agents) or thing (or set of things). To again use the analogy to parts of speech, a sentence describing an action will have an explicit subject (e.g. NASA created this video). Occurrences that are not actions, either do not have a clearly identifiable subject (e.g. It rained) or have such a diffuse set of implied subjects that it does not make sense to identify them explicitly (e.g. the Civil War). An intentional action is an action initiated by an agent (e.g. Booth's assassination of Lincoln), whereas other actions could be carried out by things (e.g. a server dynamically generating a thumbnail from a master image). An action can be transitive or intransitive. A transitive action is action upon something/someone, e.g. "Fred bought a book." An intransitive action takes no specific objects, e.g. "Fred laughed." A transitive action is performed by one entity upon another entity, rather than upon the properties of an entity (Farradane, 1980). A speech act is one type of action: a communicative event, whose meaning and appropriateness can depend significantly on the circumstances in which it takes place (Austin, 1962). When conducted through fixed symbolic forms – such as electronic mail, telephone, video – speech acts can yield persistent objects; speech acts that are carried out in face-to-face verbal conversation, on the other hand, are events that generally do not leave persistent documentary traces as evidence that they occurred.

Two general categories of occurrences that are relevant to digital collections are:

(1) those in the world that might be related to the objects preserved (e.g. death of a former president); and

(2) things that happen to the objects themselves (e.g. transformation from one file format to another, transfer of custody, use or annotation).

It is often particularly useful to identify events associated with "beginning of existence" and "end of existence" of objects and agents (ISO 21127, 2006).

5.2.4 Purpose. The purpose category includes mandates, norms, values, intentions, rules, standards, virtues, or functions which agents can:

- advance to or conform with;
- attempt to advance to or conform with;
- hope to advance to or conform with; or
- perceive/expect entities (or sets of entities) to advance to or conform with[14].

A purpose often serves as a good answer to a "why" question (e.g. Why did you create that digital object? Why did the repository take custody of it?). Purpose within a "context is a criterion in settling the question of why a man who has just put a cigarette in his mouth has put his hand in his pocket; relevance to an obvious end is a criterion in settling why a man is running away from a bull" (Grice, 1989). This category includes mandates, which are "associated with" and "govern the relationships between" other entities (Acland *et al.*, 1999), and rights, which are "standard[s] of permitted and forbidden action within a certain sphere" (*Oxford English Dictionary*, 1989). Purpose is always the purpose of agents. In some cases, one can easily relate a purpose to a specific agent (e.g. Anne sent out the meeting announcement, because she wanted her staff to attend), while other purposes are far too broad to be connected to a specific agent or set of agents (e.g. the pursuit of happiness, accountability to the citizenry).

Within social structures, purposes are often formally enacted through functions; one often addresses and pursues the function itself, rather than directly referencing the purposes it is enacting (i.e. within a given social structure, the function effectively acts as the purpose). Functions often have nested relationships with other functions, activities or transactions (Schellenberg, 1956; Robinson, 1997). The associations between functions and specific individuals or organizational units are often messy, dynamic and difficult to identify, due to:

- the amorphous and often tacit boundaries between both functions and organizational structures;
- shared responsibilities at a given point in time; and
- changes in responsibility over time.

The Australian "series system" for archival description was designed to address such mapping issues (Scott, 1966; Cunningham, 2007).

Objects cannot have purposes, but they often embody, enact and facilitate purposes. This can include the use of artifacts to fulfill specific purposes for which they were consciously designed (e.g. using a crash bar to push open a door) or more improvisational enactment of purposes through objects (e.g. using a book to hold open a door). The purposes to be fulfilled by objects are never strictly determined and can change significantly over time through acts of "reinvention" (Rogers, 1995) or "transformation" (Orlikowski, 2000) by those who use and interact with the objects. However, objects do often have affordances (Gibson, 1979) or perceived affordances (Norman, 1990) that make some purposes much easier to advance than others over time. Creators of objects can both exploit and contribute to their affordances in ways that are likely to support the creators' purposes. Objects often embody deep values and politics (Winner, 1986; Latour, 1992; Lessig, 2006). Complex digital objects also tend to reflect the functional and communication structure of the organizations or groups that

created them (Conway, 1968; Souza *et al.*, 2005). Such embedded normative, political and organizational purposes are often not apparent simply from interacting with the objects, and they will be much less apparent to users of the future. In such cases, it can be essential for the curator of a collection of objects to provide additional contextual information about "hidden" purposes, helping future users to engage in what Bowker and Star (1999) call "infrastructural inversion" and what DeSanctis and Poole (1994) call "appropriation analysis."

Complex digital objects have the ability to reproduce relatively specific sets of behaviors. As a result, it is often possible for the intended user group of a well-designed digital object to reliably gauge and carry out the digital object's intended purpose. This interactive quality of digital objects – relatively lacking in static analog documents – is both a strength and a weakness. First, designing digital objects so that they are truly "self-explanatory artifacts" (Suchman, 1987) for a particular set of agents is extremely challenging and often fails. Second, even if function/purpose of an artifact is self-explanatory to the given set of agents, it is unlikely to be self-explanatory to other agents (or to the same agents in novel situations), which can hinder and even completely prevent any meaningful use of the digital object. This is often characterized as the "brittleness" of software. Static documents, in contrast, tend to be more robust and accommodating to placement within new purposes; although their form of expression (see below) does also strongly influence the likely purposes within which they will be understood and enacted.

5.2.5 Time. A time is "a limited stretch or space of continued existence, as the interval between two successive events or acts, or the period through which an action, condition, or state continues" (*Oxford English Dictionary*, 1989). Stated more simply, it is "when something can happen." A basic distinction in database management is between "valid time" (ISO 19108, 2002), which is when some fact about the world holds true, and "transaction time," which is when the fact is reflected in the database.

The most straightforward case is a precise point in time, which can be represented as a time and date. However, the identification and application of times often involves "temporal intervals," which have varying degrees of imprecision, uncertainty, and levels of granularity (Allen, 1983). The boundary between times and relations is a matter of descriptive convention. Times are often expressed in relative terms: in relation to other specific times/occurrences (e.g. the Antebellum period) or a larger succession of times/occurrences (e.g. the 100th US Congress). When identifying times associated with the content of information artifacts (e.g. entire documents, statements within documents, fields in a database), it is important to recognize the difference between the order in which things occurred and the order in which they are reported (Sullivan *et al.*, 2008).

TIMEX2 (Ferro *et al.*, 2005) attempts to accommodate a myriad of possible temporal units and expressions, such as: decades; centuries; various ranges and durations; general terms such as past, present and future; seasons; fiscal years; sets of times (e.g. "every Tuesday"); non-specific temporal expressions (e.g. "Winters" or "on a Tuesday"); and event-anchored times (e.g. "after the death of the donor" or "the Anniversary of the bombing of Pearl Harbor").

In order to convey contextual information associated with a digital object, it can often be valuable to reflect an intersection between time and one or more of the other contextual entities. For example, it can be important to indicate what legal or other

social norms (see Purpose above) were in place at the time that something happened (event), such as the creation, reading or transfer of a digital object (Grandi *et al.*, 2005).

5.2.6 Place. A place is a designated point or region in space. Place information often can be particularly informative when associated with occurrences and agents, e.g. within biographical or organizational histories. Just a few of the many types of places are: geographic coordinates, e.g. N35-52.66 W078-47.25; relative location, e.g. next to Manning Hall; postal address, e.g. 402 E. Porter; cultural region, e.g. The South; region of national control, e.g. British Empire; city, e.g. Detroit; property or area operated by a particular institution, e.g. Goddard Space Flight Center; location in physical storage, e.g. storage bay 3, row 7, shelf 2, box 15, folder 8; and conceptual category designating control or custody, e.g. at the National Archives.

An important condition for recognizing a place can be knowing its boundaries (ISO 19107, 2003). The boundaries of places can be defined by easily recognizable physical discontinuities (e.g. a river as a boundary line) or they can be "fiat boundaries," which are invented by humans for some particular purpose (Smith and Varzi, 2000). There will often be differing views about the boundaries of a given place, or even whether a particular place exists, e.g. the eruv surrounding a Jewish community (Smith, 2007). Natural geographic features (e.g. mountains, lakes) are boundary cases that one could consider to be either objects or places; buildings similarly straddle the object-place boundary, being "not just objects, but transformations of space through objects" (Hillier and Hanson, 1984). Orientation and position are relational properties that associate the place of an object or agent with one or more other objects, agents, places or coordinate systems. Dey *et al.* (2001) use the term "location" to include not just "position information in a two-dimensional space" but also "orientation and elevation, as well as all information that can be used to deduce spatial relations between entities, such as co-location, proximity, or containment."

In many cases, there will be such a close connection between a type of place and the purposes carried out in that type of place that naming the place is almost the same thing as naming the purpose (e.g. theater, farm or drive-thru lane). However, in their discussion of "locales," Fitzpatrick *et al.* (1996) point out that place/purpose mappings are often quite complicated and dynamic. Places – especially parts of the built environment – are often closely associated with particular purposes, occurrences and relationships, as the "material preconditions for the patterns of movement, encounter and avoidance which are the material realisation – as well as sometimes the generator – of social relations" (Hillier and Hanson, 1984).

5.2.7 Form of expression. Another important aspect of a digital object's context can be whether or not it was created within some particular expressive form. This class of entities includes things about the way the digital object is expressed, which would not otherwise be evident from looking solely at the content of the digital objects itself. As early as 1876, Charles Cutter recognized "literary form" as one of the access points around which "books are most commonly brought together in catalogues" (Cutter, 1876, p. 11). This can include genre, terminology (e.g. the lingo of a particular group), and systems of measurement or exchange. This category is not intended to include technical aspects of a digital object necessary for rendering (e.g. character encoding), nor is it intended to include elements that are primarily about the content of the digital object itself.

A concept that spans both purpose and form of expression is genre. As discussed above, creators of objects often exploit affordances and perceived affordances so that the objects will advance particular purposes. Genres are sets of conventions for the creation and use of information objects that are enacted, reinforced and adapted through the ongoing flow of interactions. In short, they are "typified rhetorical actions based in recurrent situations" (Miller, 1994). As genres "become familiar, accepted, and molded through repeated use, they gain institutional force. Thus though genre emerges out of contexts, it becomes part of the context for future works" (Bazerman, 1988). Genres are not fixed, but instead "evolve over time in reciprocal interaction between institutionalized practices and individual human actions" (Yates and Orlikowski, 1992). Understanding a genre involves not only recognizing its form of expression but also appreciating relevant characteristics of the agents, occurrences, and purposes associated with the genre's origin, reinforcement and evolution.

5.2.8 Concept or abstraction. Concepts or abstractions "can be said to exist in the same sense as mathematical objects such as sets and relations, but they cannot exist at a particular place and time without some physical encoding or embodiment" (Suggested Upper Merged Ontology (SUMO), n.d.). The ABC Ontology distinguishes between an "actuality" and an abstraction (Lagoze and Hunter, 2001).

As with the boundary between target entity and context, the boundary between "real" entities (agents, occurrences, times and places) and abstractions is a subject of ongoing negotiation and evolution in almost all disciplines. Do electrons really "exist" (objects), or are they simply theoretical categories (abstractions) that help explain certain observed phenomena? Is the flow of existence inherently composed of a series of discrete events, or are events instead constructs "by which agents classify certain useful and relevant patterns of change" (Allen and Ferguson, 1994). Should one consider a fictional character to be just an idea (abstraction) or a genuine participant in a social dialog (agent)? Is the human mind reducible to a set of interconnected neurons (object), or does it have an existence apart from any particular physical instantiation (abstraction)? Should we reify (as functions) institutional and organizational structures, or should we instead think of behavior solely as an aggregation of actions and interactions (occurrences) among individual actors (agents)?

No single framework will provide definitive resolutions to the above questions. Instead, a framework for contextual information can foreground the importance, in some cases, of capturing, fixing and preserving digital representations of concepts and abstractions in order to help future users to understand, use, and make sense of TDOs. This can be particularly important in cases involving very idiosyncratic, localized or rapidly changing concepts and abstractions. What might the creator, user or intended audience of a digital object have taken a particular concept or abstraction to mean at a particular point in life of a digital object?

5.2.9 Relationship. A relationship is an association between two or more entities (or classes of entities), which cannot be reduced to or adequately expressed as a property of the entities (or classes of entities) themselves. Like an occurrence, a relationship is often transformative, i.e. the entities involved in the relationship are somehow different because of their involvement in the relationship. Some types of relationships are more closely associated with particular types of entities than others. For example, familial relationships hold between agents and not objects; relationships based on order (e.g. being before,

during or after) are more often expressed between occurrences and times than they are between places or concepts.

While it is true psychologically that "relations are not easily cognized or conceived apart from their relata" (Bliss, 1915), it is often useful to attempt such a separation within descriptions and information systems. The reason for considering relationships to be their own class of contextual entities is similar to the motivation for creating distinct "associative entities" within entity-relationship modeling in the development of databases. There are two main reasons why a relationship cannot be reduced to or expressed through the properties of the entities that are being related:

(1) the relationship can have its own attributes that are not reducible to the attributes of the entities being connected; and

(2) the independent persistence of information about the relationship can be desirable.

For example, one may want to create and retain a record of what the relationship "Facebook friend" meant at a given point in time, even if one does not currently have custody of any digital objects from Facebook that are linked using that relationship.

Within the philosophical literature, there is a long tradition of attempting to characterize the basic structures of reality using symbolic logic. The most influential approaches have often treated relations as fundamental building blocks (De Morgan, 1864; Peirce, 1880; Frege, 1917; Whitehead and Russell, 1927; Russell, 1938; Tarski, 1941; Quine, 1951). For the purposes of developing chains of inferences, two essential distinctions are between relations that are symmetrical/asymmetrical (whether or not A-relation-B implies B-relation-A) and transitive/intransitive (whether or not A-relation-B and B-relation-C implies A-relation-C).

As discussed earlier, bibliographic classification is primarily a tool for articulating assertions of aboutness. A great deal of the complexity of the classification relates to cases in which one wishes to assert that an item is about more than one thing. Faceted classification schemes are one way to address this multiple-aboutness. Each facet, as elaborated (i.e. mention of the facet) in a classification system, is designed to be independent of all the others, but any given catalog entry (i.e. use of the facets) serves as an assertion of some connection between the facets in the case of the item being described. In other words, "the combination of terms from two categories in a faceted classification implies the existence of a relation between them" (Vickery, 1960, p. 37). Vickery identified several common types of relations:

- effect;
- comparison;
- association; and
- bias.

Over several decades, J.E.L. Farradane developed a systematic approach to classification based on assertions (analets), which combine uniquely definable terms (isolates) and the relations between the terms (operators) (Farradane, 1950). He ultimately articulated nine fundamental categories of relations: concurrence, equivalence, distinctness, self-activity, dimensional, action, association, appurtenance, and functional dependence (causation) (Farradane, 1979).

As indicated earlier, relationships are essential for defining the scope and content of collections of objects. Heaney (2000) presents a detailed characterization of collection relationships. In the curation of digital data sets, a particularly important and complex set of contextual relationships are "provenance links" (Buneman *et al.*, 2006), which can help to answer "Why is a piece of data in the output?" and "Where did a piece of data come from?" (Buneman *et al.*, 2001).

A growing body of scholarship is based on the characterization and analysis of phenomena as networks in which entities are embedded, based on the insight that many relationships between entities have emergent qualities that are not visible if one only examines and determines the correlations between attributes of entities as atomized units. There is now a well-established set of methods for approaching practically any social or socio-technical phenomenon as a network of connections, interactions, or resource flows (Wasserman and Faust, 1994). In order for future users to apply such techniques (and many others that have yet to be developed) using computers, relationships must be explicitly encoded in digital form.

A role is a relationship or (more often) set of relationships between:

- one or more agents; and
- one or more purposes or occurrences.

Naming a particular role (e.g. author, custodian, mediator) tends to evoke a set of expectations about likely and appropriate behaviors and interactions on the part of whatever agent is playing that role. Much like genre expectations, role expectations can provide consistency, coherence and efficiency, but "there is always a potential for differing and sometimes conflicting expectations" (Merton, 1957), which can result in conflict, negotiation, learning and social change. In arenas of professional activity, roles are often associated with formal job titles. A role is different from (not strictly reducible to) the agent who serves in that role at a particular time and place. For example, President of the United States is a role, which is currently held by Barack Obama. This role and its related purposes preexist his term of office and they will persist after he leaves office.

6. Implications for descriptive and curatorial practice

Custodians of digital collections must decide:

- what aspects of the digital objects' creation and use environment are important enough to warrant capture, documentation, and preservation over time;
- given limited resources and available technology, which of those aspects of context can reasonably be preserved; and
- how to carry out the preservation (and often the creation and capture) of the contextual information.

6.1 Representing and preserving contextual information in digital collections

Building a network of descriptive information about the nine types of entities is different from simply trying to take a snapshot of the $context_2$ of a digital object's creation. A detailed biographic history of an author, for example, will include many pieces of information about her that were not directly present at the time she wrote a particular book (e.g. birthplace, parents' names, education, employment history). While

perhaps not directly relevant to using or making sense of a given digital object, the various biographical facts may be relevant to using or making sense of other objects that can also be associated with the biographical record. A rich set of nodes in a network of contextual information can support browsing and serendipitous discovery of such relationships.

Archival theory and practice are valuable sources of guidance for applying a contextual information framework, because they suggest that digital objects should be managed, preserved and presented in ways that reflect the social and documentary context in which they were embedded. The placement of an item within a particular category, folder or collection is "the consequence of someone's relationship with the object and, ultimately, their choice" (Currall *et al.*, 2004). Archivists attempt to reflect this form of agency by retaining the "original order" of materials. Arrangement and description of archival materials has traditionally involved identifying and gathering information related to those who "created, accumulated, assembled, or used a group of records," "function or roles records were created to support," "recordkeeping practices," and "events or developments to which the records relate" (Roe, 2005). Several authors have argued for the importance of contextual information as distinct from bibliographic information (Evans, 1986; Bearman, 1992; Roe, 1992; Thibodeau, 1995). Ideally, contextual information within a collection will not only reflect the provenance and relationships between the digital objects, but it will also shed light on the selection and description practices of the curators of collections. Hedstrom argues for "archivists to place not only the records they deal with in context – but also to place archivists, archival practice, and archival institutions in an equally dynamic context" (Hedstrom, 2002).

Descriptive standards for libraries and archives have historically been much more thorough in specifying how to describe the things within a collection than in specifying how to describe the context surrounding those things (Bearman, 1992). However, embedded within and distributed across the descriptive products of libraries and archives are numerous elements of contextual information, e.g. authority control records associated with library catalogs, and provenance, biographical and scope information within archival finding aids. Buckland (2007) explains how the types of contextual information that have traditionally been part of the library "reference collections" can be enhanced and repurposed in a networked environment.

As discussed earlier, contextual information is often characterized by recursive relationships. A component of information that was created or captured to provide contextual information about a TDO can often itself become a TDO, about which one may create, capture, and manage additional contextual information. For example, major standards that support the creation of archival finding aids – Encoded Archival Description (EAD) (Society of American Archivists and Library of Congress, 2002); General International Standard Archival Description (ISAD(G)) (*Ad Hoc* Commission on Descriptive Standards, 2000); Rules for Archival Description (RAD) (Canadian Council of Archives, 2003); and Describing Archives: A Context Standard (DACS) (Society of American Archivists, 2004) – include elements related to the context of creation of the finding aids themselves. Light and Hyry (2002) argue for the incorporation of further contextual information throughout the lifecycle of the finding aid.

One of the complications of digital objects is that they do not reflect the same properties in all contexts – i.e. they do not conform to Leibniz's Law (Allison *et al.*,

2004). First, at a given time, the same digital object might have different properties when rendered and used in two different computing environments. Second, the preservation of a digital object over time can involve transforming various components – and as a result, the properties and behaviors – of the digital object.

There is not one a priori way of determining whether a given instantiation represents transformation into a new digital object. One may either consider each distinct instance of the digital object to be a new item worthy of its own identity and management, or all of the separate instances to be enactments of the same digital object, but in different contexts. When a new instance of the digital object represents a transaction about which one hopes to have evidence (including, e.g. how the digital object was perceived), then option 1 can be desirable. When one's primary concern is the full life (including provenance and chain of custody) of a particular logical entity, such as an official government record, then it can be useful to apply option 2. For any given interaction one has with a TDO, one could treat the TDO as a new work, new expression of an existing work, new manifestation of an existing expression, new item of an existing manifestation (IFLA Study Group, 1988), or unique projection onto computer hardware of an existing item (e.g. identical bitstream mirrored or cached in a second location).

6.2 Capturing contextual information over time: approaches, scope and limits

Like a snowball rolling down a hill, which picks up a subset of the snow that it touches along the way, TDOs can pick up pieces of contextual information over time. Figure 2 presents creation, capture and preservation of contextual information, within a set of broad stages developed in the DigCCurr project (Lee *et al.*, 2007). Reference collections within repositories can serve as sources of contextual information (Buckland, 2007), as well as supplemental sources created outside repositories, such as directories and "the vast compilation of data by genealogists" (Hurley, 1995). There are also potential opportunities throughout (and even before and after) the existence of a TDO to add contextual information. Some important contextual information is best created or captured before the TDO has been created – e.g. documenting administrative processes that generate institutional records; describing the methodology, instruments, and procedures involved in collecting research data. The variability across studies in ecology, for example, "necessitates that contextual information for ecological data are carefully recorded starting with the data planning and subsequently with the actual data taking and data curation" (Karasti *et al.*, 2006). It can also be desirable to create and capture contextual information at or close to the time of a digital object's creation, by those who are engaged in the activities that generate the TDOs (Hedstrom, 1993; Wallace, 1995; McDonald, 1997).

The generation of important contextual information often does not stop at the point of creation. If metadata is assigned to a document at only one point in time, then "one cannot anticipate interpretations or usefulness of a document in the light of later knowledge or viewpoints" (Farradane, 1980). Stated another way, curators of digital collections can capture and preserve not only the primary (original) context of digital objects, but also their "secondary" context or provenance (Sharer and Ashmore, 1987). "A document is more than its subject content and the context of its creation. Throughout its life cycle, it continually evolves, acquiring additional meanings and layers, even after crossing the archival threshold." Ongoing curation of contextual

information should go "beyond the 'snapshot of information'" found in a static descriptive tool such as a finding aid (Nordland, 2004). One example of such contextual information is the "publication note" in ISAD(G), which is used "To identify any publications that are about or are based on the use, study, or analysis of the unit of description." The "Bibliography" element in EAD is more broadly reserved for "works that are based on, about, or of special value when using the materials being described, or works in which a citation to or brief description of the materials is available." The PREMIS Data Dictionary represents an important step in foregrounding the importance of and explicating ways to capture and represent information about changes to digital objects over time.

There are practical limitations to how much contextual information can be captured, and there are also reasons why not all contextual information should be captured, even if this were possible. These include issues of professional ethics (e.g. confidentiality, cultural property) and usability. The reuse of a digital object requires "lifting" (Guha and McCarthy, 2003) it out of its original context and then making sense and use of it in a new context. Support for such lifting requires a balance between providing too little contextual information, so that the user does not understand what she is interacting with, and too much contextual information, so that she "will drown in unnecessary, unhelpful, or conflicting data" (Ackerman and Halverson, 1998). A similar finding from the development of expert systems is that it is "highly unlikely that an expert will provide a rule which is completely wrong under all circumstances; [instead] the more likely problems are excessive generalisation, or attention to peripheral data" (Compton and Jansen, 1990).

Existing approaches for creating and capturing contextual information associated with TDOs within collections have usually attended to particular, relatively discrete points in the information lifecycle, for example: creation; declaration or endorsement; publication; and transfer into the repository. Even this fairly limited set of points in the lifecycle can become cumbersome to document in any great detail, when this depends on the attention and direct intervention of professional curators.

Another potential source of contextual information is use data, particularly if it is "collected from users in an ongoing program" (Conway, 1986). Many repositories collect information about the use of their materials, but ethical and administrative issues have often hindered their ability to retain the use data for long periods of time. Even when institutions do have many years of use data, it often does not provide access points that allow curators or end users to associate it easily with specific objects. There is also wide variance in the forms of use data collected, which makes it very challenging to analyze, exchange or compare.

Users of digital collections can also identify, capture and create contextual information. There will always be limits to how much potentially relevant contextual information for a given digital object a curator can determine *a priori*. The notion of context is inherently "interactional" (Shanon, 1990). Therefore, one important consideration is how the archive might identify and ingest elements of contextual information that emerge in the environment (e.g. users creating "artificial collections" that tie objects together in unpredicted ways, annotations, new derived surrogates, analytical tools). A "scholar may have far greater understanding of the content and context of a particular artifact, and/or greater sophistication about the implications of specific approaches to *re-presentation*" (Bearman and Trant, 1998). Some users will be

able to bring unique information to the archive about aspects of the $context_1$, $context_2$, and $context_3$ of a digital object from stages in its lifecycle before it was ingested by the archive.

Users can also contribute contextual information about points in the life of a digital object after it has been transferred from its original use environment and into the archive. Throughout "the life of an information resource people and organizations playing quite different roles will create metadata that may be germane to future users" (Bearman and Trant, 1999). Recent developments in the Semantic Web and social tagging hold great promise for supporting the capture and preservation of contextual information. Users can provide annotations, overlays and other value-added elements to digital objects, which reflect characteristics of their own user experiences and what sense they have made of the digital objects. As represented in the life cycle model of the Data Documentation Initiative, various forms of data set "repurposing" – e.g. "streamlined instructional data set, a specific sampling and restructuring of the data, or combining data from multiple sources to create a new data set (either physically or virtually)" – can also be fed back into an archive for long-term preservation (Thomas *et al.*, 2008).

There is great promise in using automated or semi-automated techniques to generate contextual information, including the following (all of which have been addressed by recent research):

- named entity recognition or metadata extraction based on machine learning, natural language processing or recognition of elements embedded into particular files;
- developing collections of facts about known entities using data available on the web;
- inferring the authors of portions of text within a collection;
- using patterns in the co-occurrence of dates and places within documents to identify events within a collection;
- machine learning to make inferences about temporal relations between events;
- natural language processing to generate suggestions about who authored a particular digital object and intended sentiment or emotional state of the author; and
- network analysis to determine degree of social influence of a particular agent.

Such inferred contextual information (potentially confirmed, further described or selected by a human curator) could then be ingested and associated with the appropriate digital objects. The elements of contextual information most likely to require direct capture or creation by human curators are those that are discrete, localized and ephemeral, thus not well-documented elsewhere, subtle or complex enough that they would be difficult to identify or disambiguate using automation later, fundamental bridging mechanisms to collections that reside elsewhere, or so fundamental to the collecting mission of the archive that they warrant focused, narrative description.

Preservation of contextual information does not always require direct custody. A repository that has TDOs for which it hopes to provide further contextual information can enter into cooperating, federated, or shared resource (OAIS) arrangements with

other trustworthy archives, who are responsible for managing the contextual information. For purposes of access, a repository can provide links to context items within other repositories, either directly or through a federated access system. It is important to note, however, that most arrangements for federating collections to date have involved extraction of discrete items or metadata elements for purposes of discovery and retrieval. While "federated collections can conceivably be built to offer more than the sum of their parts, these aggregations may also lose important context and meaning inherent in individual collections" (Palmer *et al.*, 2007). Some forms of contextual information can also be presented to users on-the-fly through "context interpreter" software that maps from one context item to another based on known associations, e.g. presenting the email address associated with a name (Dey, 2001). By directly attending to the creation, capture, management and sharing of contextual information, curators of digital collections can best ensure that the distributed network of digital collections will provide not only access to digital objects but also the means to make meaningful use and sense of the digital objects long into the future.

Conclusion

Information professionals have long recognized the importance of providing contextual information to users of the items in their care. The representation of contextual information has been a core element of the theory and practice of archival description for centuries, and it is receiving increasing attention in the literature on digital libraries and digital preservation. However, there has previously been a detailed elaboration of what it means to provide contextual information within a digital collection. This paper has provided a framework of contextual information that is based on a synthesis of a diverse and extensive body of literature about context and professional guidance documents. I believe this represents an important step in the ongoing evolution of thinking and practice in the curation of digital collections.

Notes

1. An earlier articulation of these ideas is available as: Lee (2007).
2. One formulation of the "context" of a statement is precisely that portion of $context_2$ that is not reflected in $context_1$. That is, context is "an abstraction of the features that are not explicitly included" in a model of the world (Edmonds, 1999). This type of context, by definition, will never be captured or preserved explicitly within an information system.
3. An important aspect of context can be the culture in which a TE is embedded or enacted. As a set of shared patterns of behavior (social structure), culture is often best characterized as a form of $context_2$. However, any culture worthy of the name will also manifest itself in various ways through $context_1$ and $context_3$.
4. To the extent that they serve as "representational artifacts" (Levy, 2001), surrounding objects can serve as both $context_2$ and $context_1$. The arrangement of documents relative to each other within a physical space (or experientially physical but digitally mediated space such as a computer file system or desktop) can constitute documentary context, but it can also shape habitual and embodied behavior in ways that do not involve direct symbolic processing (Malone, 1983; Kirsh, 2001; Sellen and Harper, 2002).
5. The latter is an essential factor in game theory, social psychology, sociolinguistics and social epistemology, e.g. whether A can safely assume that a person with whom she is interacting has particular preferences, expectations or acquaintance with facts.

6. Studies of the role of framing within social phenomena often lie at the intersection of $context_1$ and $context_3$. Predominant questions are both how individuals interpret or act upon messages within their own lives, and how this personal experience aligns (or fails to align) with the experience of others, which is often influenced by a message's placement within a shared discourse ($context_1$).
7. A theory or other system of inter-related background statements/assumptions is a case of context1 when it is explicitly encoded into an information system (see, e.g. McCarthy, 1993; Guha, 1991; Guha and Lenat, 1994), but it can constitute part of $context_3$ when it is "in the head" of a particular agent or is otherwise brought to bear on her interpretations of, understanding of, actions within or noticing of features within the situation at hand.
8. Many other entities that have played a part in the full life history of books also appear in catalog entries, but usually receive only peripheral attention. For example, the Catalog of Printed Books in the British Museum from 1841 provides a hint at the potential importance of chain of custody information, but limits this to a "few words at the end of the entry" in cases when "the volume belonged to some very distinguished personage" (Panizzi, 1841). Contemporary cataloging standards include fields in which the circumstance of creation and some chain of custody information can be articulated, e.g. Date/Time and Place of an Event Note (field 518), Immediate Source of Acquisition Note (field 541), Ownership and Custodial History (field 561), and Accumulation and Frequency of Use Note (field 584) in MARC 21 (US Library of Congress, 2010). However, entities with relationships to an item after or before it has been created, enacted or performed are not usually emphasized in bibliographic classification systems; they are often documented in the descriptions of special collections and archives, which are discussed below.
9. Applying traditional library cataloging principles can be particularly challenging with non-bibliographic materials. Analyzing the subject of a picture, for example, can involve both what the picture is about and what the picture is of, at multiple levels of meaning (Shatford, 1986).
10. As discussed earlier, an aggregation of fixed representations, no matter of detailed, will never fully reflect all aspects of the reality through which a digital object has passed over time. However, "comprehensive" documentation is a useful hypothetical limiting case for purposes of elaborating a model or framework.
11. Persistent state information is information that is stored in order to be accessed across computer sessions (Rajasekar *et al.*, 2010), however a session may be defined. For example, persistent state information could help an individual in "relating current browsing for PC accessories to last year's purchase of a desktop machine" (Kramer *et al.*, 2000).
12. See, for example, the distinctions in FRBR between work, expression, manifestation and item; and in PREMIS between intellectual entities and digital objects.
13. A closely related matter, which has fluctuated among scholars over time, is which types of entities (e.g. humans, all animate beings, computerized devices, all physical objects) have a capacity for "interaction" (Suchman, 1987).
14. I was not able to identify an English word that fully reflects this category of "whyness." In using the term "purpose," I do not intend to restrict this category to a rational-choice formulation of purpose as the attainment of a well-defined end or end state.

References

Ackerman, M.S. and Halverson, C. (1998), "Considering an organization's memory", *CSCW '98: Proceedings: ACM 1998 Conference on Computer Supported Cooperative Work, Seattle, Washington, November 14-18*, Association for Computing Machinery, New York, NY, pp. 39-48.

Acland, G., Cumming, K. and McKemmish, S. (1999), "The end of the beginning: the SPIRT Recordkeeping Metadata Project", paper presented at the Australian Society of Archivists Conference, Brisbane, Australia, 29-31 July 1999, available at: www.archivists.org.au/events/conf99/spirt.html (accessed 1 September 2009).

Ad Hoc Commission on Descriptive Standards (2000), *ISAD(G): General International Standard Archival Description*, 2nd ed., International Council on Archives, Ottawa.

Adams, D. (1980), *The Restaurant at the End of the Universe*, Pocket Books, New York, NY.

Agre, P.E. (2001), "Changing places: contexts of awareness in computing", *Human-Computer Interaction*, Vol. 16 Nos 2-4, pp. 177-92.

Alexandria Digital Library Feature Type Thesaurus (2002), University of California, Santa Barbara, CA, available at: www.alexandria.ucsb.edu/gazetteer/FeatureTypes/ver070302/ (accessed 1 September 2009).

Alexandria Digital Library Project (2004), "Guide to the ADL Gazetteer Content Standard", Version 3.2, Alexandria Digital Library Project, Santa Barbara, CA, available at: www.alexandria.ucsb.edu/gazetteer/ContentStandard/version3.2/GCS3.2-guide.htm (accessed 29 August 2009).

Allen, J.F. (1983), "Maintaining knowledge about temporal intervals", *Communications of the ACM*, Vol. 26 No. 1, pp. 832-43.

Allen, J.F. (1984), "Towards a general theory of action and time", *Artificial Intelligence*, Vol. 23 No. 2, pp. 123-54.

Allen, J.F. and Ferguson, G. (1994), "Actions and events in interval temporal logic", *Journal of Logic and Computation*, Vol. 4 No. 5, pp. 531-79.

Allison, A., Currall, J., Moss, M. and Stuart, S. (2004), "Digital identity matters", *Journal of the American Society for Information Science and Technology*, Vol. 56 No. 4, pp. 364-72.

American Library Association (2008), "Resource Description and Access (RDA)", American Library Association, Canadian Library Association, and Chartered Institute of Library and Information Professionals, 2008, available at: www.rdaonline.org/constituencyreview/ (accessed April 12, 2010).

Austin, J.L. (1962), *How to Do Things with Words*, Harvard University Press, Cambridge, MA.

Australian Government Recordkeeping Metadata Standard (2008), "Version 2.0", National Archives of Australia, available at: www.naa.gov.au/records-management/publications/AGRkMS.aspx (accessed 30 August 2009).

Australian Governments' Interactive Functions Thesaurus – AGIFT (2007), 2nd ed., National Archives of Australia, Canberra.

Baca, M. and Harpring, P. (Eds) (2009), "Categories for the description of works of art", J. Paul Getty Trust, Los Angeles, CA, available at: www.getty.edu/research/conducting_research/standards/cdwa/ (accessed 1 September 2009).

Bardram, J.E. (2005), "The Java context awareness framework (JCAF) – a service infrastructure and programming framework for context-aware applications", in Gellersen, H., Want, R. and Schmidt, A. (Eds), *Pervasive Computing: 3rd International Conference, PERVASIVE 2005, Munich, Germany, May 8-13, 2005, Proceedings,* Springer, Berlin, pp. 98-115.

Bates, M.J. (1996), "The Getty End-user Online Searching Project in the Humanities: Report No. 6: Overview and Conclusions", *College & Research Libraries*, Vol. 57, pp. 514-23.

Bates, M.J., Wilde, D.N. and Siegfried, S. (1993), "An analysis of search terminology used by humanities scholars: the Getty Online Searching Project Report Number 1", *Library Quarterly*, Vol. 63 No. 1, pp. 1-39.

Bazerman, C. (1988), *Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science*, University of Wisconsin Press, Madison, WI.

Bearman, D. (1989), *Archival Methods*, Archives and Museum Informatics, Pittsburgh, PA.

Bearman, D. (1992), "Documenting documentation", *Archivaria*, Vol. 34, pp. 33-49.

Bearman, D. and Trant, J. (1998), "Authenticity of digital resources: towards a statement of requirements in the research process", *D-Lib Magazine*, Vol. 4 No. 6.

Bearman, D. and Trant, J. (1999), "Unifying our cultural memory: could electronic environments bridge the historical accidents that fragment cultural collections?", in Criddle, S., Dempsey, L. and Heseltine, R. (Eds), *Information Landscapes for a Learning Society: Networking and the Future of Libraries 3*, Library Association Publishing, London, pp. 207-34.

Bearman, D.A. and Lytle, R.H. (1985), "The power of the principle of provenance", *Archivaria*, Vol. 21, pp. 14-27.

Bekaert, J.L., De Kooning, E. and Van de Walle, R. (2005), "Packaging models for the storage and distribution of complex digital objects in archival information systems: a review of MPEG-21 DID principles", *Multimedia Systems*, Vol. 10 No. 4, pp. 286-301.

Biographical Directory of the United States Congress (n.d.), available at: http://bioguide.congress.gov/ (accessed 1 September 2009).

Birnholtz, J.P. and Bietz, M.J. (2003), "Data at work: supporting sharing in science and engineering", in Prendergast, M. (Ed.), *Group '03: Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work: November 9-12, 2003, Sundial Resort on Sanibel Island, Florida, USA*, Association for Computing Machinery, New York, NY, pp. 339-48.

Bliss, H.E. (1915), "On relations", *Philosophical Review*, Vol. 24 No. 1, pp. 37-53.

Blum, A. (2003), *The Imaginative Structure of the City*, McGill-Queen's University Press, Montreal.

Bock, C. and Gruninger, M. (2005), "PSL: a semantic domain for flow models", *Software and Systems Modeling*, Vol. 4 No. 2, pp. 209-31.

Bourdieu, P. (1977), *Outline of a Theory of Practice*, Cambridge University Press, Cambridge, (R. Nice, trans.).

Bowker, G.C. and Star, S.L. (1999), *Sorting Things Out: Classification and its Consequences*, MIT Press, Cambridge, MA.

Brack, E.V., Palmer, D. and Robinson, B. (2000), "Collection level description – the RIDING and Agora experience", *D-Lib Magazine*, Vol. 6 No. 9.

Brézillon, P. (1999), "Context in problem solving: a survey", *Knowledge Engineering Review*, Vol. 14 No. 1, pp. 47-80.

Brickley, D. and Miller, L. (2007), *FOAF Vocabulary Specification 0.91*, available at: http://xmlns.com/foaf/spec/ (accessed 29 August 2009).

Broughton, V. (2010), "Bliss bibliographic classification: 2nd edition", in Bates, M.J. and Maack, M.N. (Eds), *Encyclopedia of Library and Information Sciences*, CRC Press, Boca Raton, FL, pp. 650-9.

Buchanan, G., Cunningham, S.J., Blandford, A., Rimmer, J. and Warwick, C. (2005), "Information seeking by humanities scholars", in Christodoulakis, S., Rauber, A. and Tjoa, A.M. (Eds), *Research and Advanced Technology for Digital Libraries, 9th European Conference, ECDL 2005, Vienna, Austria, September 18-23, 2005: Proceedings*, Springer, Berlin, pp. 218-29.

Buckland, M.K. (2007), "The digital difference in reference collections", *Journal of Library Administration*, Vol. 46 No. 2, pp. 87-100.

Buneman, P., Chapman, A. and Cheney, J. (2006), "Provenance management in curated databases", *ACM SIGMOD/PODS 2006 Electronic Proceedings: Chicago, Illinois, USA, June 26-29, 2006*, Association for Computing Machinery, New York, NY, pp. 539-50.

Buneman, P., Khanna, S. and Tan, W.-C. (2001), "Why and where: a characterization of data provenance", in Bussche, J.V.D. and Vianu, V. (Eds), *Database Theory – ICDT 2001: 8th International Conference, London, UK, January 2001, Proceedings*, Springer, Berlin, pp. 316-30.

Callon, M. and Law, J. (1989), "On the construction of sociotechnical networks: content and context revisited", *Knowledge and Society*, Vol. 8, pp. 57-83.

Canadian Council of Archives (2003), *Rules for Archival Description*, Canadian Council of Archives, Ottawa.

Çelik, T. (2007), "Geo-Microformats", available at: http://microformats.org/wiki/geo (accessed 29 August 2009).

Chen, P.P-S. (1976), "The entity-relationship model – toward a unified view of data", *ACM Transactions on Database Systems*, Vol. 1 No. 1, pp. 9-36.

Chin, G. Jr and Lansing, C.S. (2004), "Capturing and supporting contexts for scientific data sharing via the biological sciences collaboratory", *CSCW 2004: Computer Supported Cooperative Work: Conference Proceedings, Chicago, IL, November 6-10, 2004*, ACM Press, New York, NY, pp. 409-18.

Clark, H.H. (1996), *Using Language*, Cambridge University Press, Cambridge.

Cole, C. (2000), "Name collection by PhD history students: inducing expertise", *Journal of the American Society for Information Science*, Vol. 51 No. 5, pp. 444-55.

Compact Oxford English Dictionary of Current English (2005), 3rd ed., Oxford University Press, New York, NY.

Compton, P. and Jansen, R. (1990), "Knowledge in context: a strategy for expert system maintenance", paper presented at the 2nd Australian Joint Conference on Artificial Intelligence, Adelaide, Australia.

Consultative Committee for Space Data Systems (2008), "XML formatted data unit (XFDU) structure and construction rules", Consultative Committee for Space Data Systems, Washington, DC, Draft recommendation for space data system standards.

Conway, M.E. (1968), "How do committees invent?", *Datamation*, Vol. 14 No. 4, pp. 28-31.

Conway, P. (1986), "Facts and frameworks: an approach to studying the users of archives", *American Archivist*, Vol. 49 No. 4, pp. 393-407.

COUNTER – Online Usage of Electronic Resources (n.d.), available at: http://projectcounter.org/ (accessed 30 August 2009).

Crane, G. and Jones, A. (2006), "The challenge of Virginia Banks: an evaluation of named entity analysis in a 19th-century newspaper collection", *6th ACM/IEEE-CS Joint Conference on Digital Libraries 2006 Opening Information Horizons, Chapel Hill, NC, USA, June 11-15, 2006*, ACM Press, New York, NY, pp. 31-40.

Crowston, K. and Kwasnik, B.H. (2003), "Can document-genre metadata improve information access to large digital collections?", *Library Trends*, Vol. 52 No. 2, pp. 345-61.

Cunningham, A. (2007), "Harnessing the power of provenance in archival description: an Australian perspective on the development of the second edition of ISAAR(CPF)", *Journal of Archival Organization*, Vol. 5 Nos 1/2, pp. 15-31.

Currall, J., Moss, M. and Stuart, S. (2004), "What is a collection?", *Archivaria*, Vol. 58, pp. 131-46.

Cutter, C.A. (1876), "Rules for a printed dictionary catalogue", *Public Libraries in the United States of America: Their History, Condition, and Management*, US Government Printing Office, Washington, DC, pp. 9-85.

Dawson, F. and Howes, T. (1998), "vCard MIME Directory Profile" (RFC 2426), memo, Network Working Group.

Dawson, F. and Stenerson, D. (1998), "Internet calendaring and scheduling core object specification (iCalendar)" (RFC 2445), memo, Network Working Group.

De Morgan, A. (1864), "On the syllogism, No. IV, and on the logic of relations", *Transactions of the Cambridge Philosophical Society*, Vol. 10, pp. 331-58.

Dervin, B. (1983), "An overview of sense-making research: concepts, methods and results to date", paper presented at the Annual Meeting of the International Communication Association, Dallas, TX, 26-30 May, available at: www.ideals.illinois.edu/bitstream/handle/2142/2281/Dervin83a.htm (accessed 5 March 2010).

Dervin, B. (1997), "Given a context by any other name: methodological tools for taming the unruly beast", in Vakkari, P., Savolainen, R. and Dervin, B. (Eds), *Information Seeking in Context*, Taylor Graham, London, pp. 13-38.

DeSanctis, G. and Poole, M.S. (1994), "Capturing the complexity in advanced technology use: adaptive structuration theory", *Organization Science*, Vol. 5 No. 2, pp. 121-47.

Dewey, J. (1922), *Human Nature and Conduct: An Introduction to Social Psychology*, Henry Holt & Company, New York, NY.

Dewey, J. (1931), "Context and thought", *University of California Publications in Philosophy*, Vol. 12 No. 3, pp. 203-24.

Dey, A.K. (2001), "Understanding and using context", *Personal and Ubiquitous Computing*, Vol. 5, pp. 4-7.

Dey, A.K., Abowd, G.D. and Salber, D. (2001), "A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications", *Human-Computer Interaction*, Vol. 16 Nos 2-4, pp. 97-166.

Digital Library Federation (2007), "Metadata encoding and transmission standard: primer and reference manual", Digital Library Federation, Washington, DC, available at: www.loc.gov/standards/mets/METS%20Documentation%20final%20070930%20msw.pdf (accessed 1 September 2009).

Doerr, M. and Kritsotaki, A. (2006), "Documenting events in metadata", paper presented at the 7th International Symposium on Virtual Reality, Archaeology and Cultural Heritage, Nicosia, Cyprus, 30 October-4 November.

Dourish, P. (2001), *Where the Action Is: The Foundations of Embodied Interaction*, MIT Press, Cambridge, MA.

Duff, W. and Harris, V. (2002), "Stories and names: archival description as narrating records and constructing meanings", *Archival Science*, Vol. 2 Nos 3-4, pp. 263-85.

Duff, W.M. and Johnson, C.A. (2001), "A virtual expression of need: an analysis of e-mail reference questions", *American Archivist*, Vol. 64 No. 1, pp. 43-60.

Edmonds, B. (1999), "The pragmatic roots of context", in Bouquet, P., Serafini, L., Brezillon, P., Benerecetti, M. and Castellani, F. (Eds), *Modeling and Using Context: 2nd International and Interdisciplinary Conference, CONTEXT'99, Trento, Italy, September 1999, Proceedings*, Springer-Verlag, Berlin, pp. 119-34.

Evans, M.J. (1986), "Authority control: an alternative to the record group concept", *American Archivist*, Vol. 49, pp. 249-61.

Farradane, J. (1979), "Relational indexing. Part I", *Journal of Information Science*, Vol. 1 No. 5, pp. 267-76.

Farradane, J. (1980), "Relational indexing. Part II", *Journal of Information Science*, Vol. 1, pp. 313-24.

Farradane, J.E.L. (1950), "A scientific theory of classification and indexing and its practical applications", *Journal of Documentation*, Vol. 6 No. 2, pp. 83-99.

Ferro, L., Gerber, L., Mani, I., Sundheim, B. and Wilson, G. (2005), "TIDES 2005 standard for the annotation of temporal expressions", MITRE Corporation, available at: http://timex2.mitre.org/annotation_guidelines/2005_timex2_standard_v1.1.pdf (accessed 1 September 2009).

Fillmore, C.J. (1977), "Scenes-and-frames semantics", in Zampolli, A. (Ed.), *Linguistic Structures Processing*, North-Holland, Amsterdam, pp. 55-81.

Fitzpatrick, G., Kaplan, S. and Mansfield, T. (1996), "Physical spaces, virtual places and social worlds: a study of work in the virtual", *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, ACM Press, New York, NY, pp. 334-43.

Foot, K., Schneider, S. and Dougherty, M. (2003), "The ethics of/in web archiving", paper presented at Broadening the Band: Association of Internet Researchers Conference, Toronto, Ontario, 16-19 October, available at: www.webarchivist.org/031009.ethics-web-archiving_ReadOnly.ppt (accessed 1 September 2009).

Franklin, B. and Plum, T. (2006), "Successful web survey methodologies for measuring the impact of networked electronic services (MINES for Libraries)", *IFLA Journal*, Vol. 32 No. 1, pp. 28-40.

Frege, G. (1917), "Class, function, concept, relation", *The Monist*, Vol. 27, pp. 114-27.

Gadamer, H.G. (1989), *Truth and Method*, 2nd rev. ed., translated by Weinsheimer, J. and Marshall, D.G., Crossroad, New York, NY.

Gibson, J.J. (1979), *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston, MA.

Glaser, B.G. and Strauss, A.L. (1964), "Awareness contexts and social interaction", *American Sociological Review*, Vol. 29 No. 5, pp. 669-79.

Goffman, E. (1959), *The Presentation of Self in Everyday Life*, Doubleday, Garden City, NY.

Goffman, E. (1974), *Frame Analysis: An Essay on the Organization of Experience*, Harvard University Press, Cambridge, MA.

Grandi, F., Mandreoli, F. and Tiberio, P. (2005), "Temporal modelling and management of normative documents in XML format", *Data & Knowledge Engineering*, Vol. 54 No. 3, pp. 327-54.

Greenberg, S. (2001), "Context as a dynamic construct", *Human-Computer Interaction*, Vol. 16 Nos 2/4, pp. 257-68.

Grice, H.P. (1989), *Studies in the Way of Words*, Harvard University Press, Cambridge, MA.

Grudin, J. (2001), "Desituating action: digital representation of context", *Human-Computer Interaction*, Vol. 16 Nos 2/4, pp. 269-86.

GS1 US (2006), "An introduction to the global trade item number (GTIN)", Lawrenceville, NJ, available at: http://barcodes.gs1us.org/dnn_bcec/Documents/tabid/136/DMXModule/731/Command/Core_Download/Default.aspx?EntryId=59 (accessed 1 September 2009).

Guha, R.V. (1991), "Contexts: a formalization and some applications", PhD dissertation, Stanford University, Stanford, CA.

Guha, R.V. and Lenat, D.B. (1994), "Enabling agents to work together", *Communications of the ACM*, Vol. 37 No. 7, pp. 126-42.

Guha, R. and McCarthy, J. (2003), "Varieties of vontexts", in Blackburn, P., Ghidini, C., Turner, R.M. and Giunchiglia, F. (Eds), *Modeling and Using Context: 4th International and*

Interdisciplinary Conference CONTEXT 2003 Stanford, CA, USA, June 23-25, 2003 Proceedings, Springer, Berlin, pp. 164-77.

Harpring, P., Beecroft, A., Johnson, R. and Ward, J. (Eds) (2006), *Union List of Artist Names: Editorial Guidelines*, J. Paul Getty Trust, Los Angeles, CA.

Heaney, M. (2000), "An analytical model of collections and their catalogues, third issue", UK Office for Library and Information Networking, available at: www.ukoln.ac.uk/metadata/rslp/model/amcc-v31.pdf (accessed 1 September 2009).

Heckmann, D. and Krüger, A. (2003), "A user modeling markup language (UserML) for ubiquitous computing", in Brusilovsky, P., Corbett, A. and de Rosis, F. (Eds), *User Modeling 2003: 9th International Conference, UM 2003 Johnstown, PA, USA, June 22-26, 2003 Proceedings*, Springer, Heidelberg, pp. 393-7.

Heckmann, D., Schwartz, T., Brandherm, B., Schmitz, M. and von Wilamowitz-Moellendorff, M. (2005), "GUMO: the general user model ontology", in Ardissono, L., Brna, P. and Mitrovic, A. (Eds), *User Modeling 2005: 10th International Conference, UM 2005, Edinburgh, Scotland, July 24-29, 2005: Proceedings*, Springer, Berlin, pp. 428-32.

Hedstrom, M. (1993), "Descriptive practices for electronic records: deciding what is essential and imagining what is possible", *Archivaria*, Vol. 36, pp. 53-63.

Hedstrom, M. (2002), "Archives, memory, and interfaces with the past", *Archival Science*, Vol. 2 Nos 1/2, pp. 41-43.

Hedstrom, M. and King, J.L. (2006), "Epistemic infrastructure in the rise of the knowledge economy", in Kahin, B. and Foray, D. (Eds), *Advancing Knowledge and the Knowledge Economy*, MIT Press, Cambridge, MA, pp. 113-34.

Hedstrom, M. and Lee, C.A. (2002), "Significant properties of digital objects: definitions, applications, implications", *Proceedings of the DLM-Forum 2002, Barcelona, 6-8 May 2002: @ccess and Preservation of Electronic Information: Best Practices and Solutions*, Office for Official Publications of the European Communities, Luxembourg, pp. 218-27.

Hedstrom, M.L., Lee, C.A., Olson, J.S. and Lampe, C.A. (2006), "'The old version flickers more': digital preservation from the user's perspective", *American Archivist*, Vol. 69 No. 1, pp. 159-87.

Heidegger, M. (1996), *Being and Time: A Translation of Sien und Zeit*, translated by Stambaugh, J., State University of New York Press, Albany, NY.

Henderson, S. (2005), "Genre, task, topic and time: facets of personal digital document management", in Plimmer, B. and Amor, R. (Eds), *CHINZ 2005 – Making CHI Natural: Proceedings: 6th International Conference NZ Chapter of the ACM's Special Interest Group on Computer-Human Interaction (SIGCHI-NZ), 7-8 July 2005 Auckland, New Zealand*, Association for Computing Machinery, New York, NY.

Hillier, B. and Hanson, J. (1984), *The Social Logic of Space*, Cambridge University Press, Cambridge.

Hjørland, B. and Albrechtsen, H. (1995), "Toward a new horizon in information science: domain-analysis", *Journal of the American Society for Information Science*, Vol. 46 No. 6, pp. 400-25.

Holdsworth, D. and Wheatley, P. (2004), "Long-term stewardship of globally-distributed representation information", paper presented at the 12th NASA Goddard/21st IEEE Conference on Mass Storage Systems and Technologies: Long-term Stewardship of Globally Distributed Storage, Adelphi, MD, 13-16 April, available at: http://storageconference.org/STORAGECONFERENCE/2004/Papers/03-Holdsworth-a.pdf (accessed 1 September 2009).

Hoyt, J., Daugherty, J. and Recordon, D. (2006), "OpenID Simple Registration Extension 1.0", available at: http://openid.net/specs/openid-simple-registration-extension-1_0.html (accessed 16 July 2007).

Hurley, C. (1995), "Problems with provenance", *Archives and Manuscripts*, Vol. 23 No. 2, pp. 234-59.

Husserl, E. (1952), *Ideas: General Introduction to Pure Phenomenology*, Allen & Unwin, London, (W.R.B. Gibson, Trans.).

ICA Committee on Best Practices and Standards (2008), *ISDF: International Standard for Describing Functions*, 1st ed., International Council on Archives, Paris.

ICA Committee on Descriptive Standards (2004), *International Standard Archival Authority Record for Corporate Bodies, Persons and Families (ISAAR/CPF)*, 2nd ed., International Council on Archives, Paris.

IDEAlliance (2006), *IDEAlliance Publishing Requirements for Industry Standard Metadata: Guide to the PRISM Digital Image Management Metadata Encoding*, International Digital Enterprise Alliance, Alexandria, VA, Version 1.0.

IFLA (1998), "Functional requirements for bibliographic records: final report", IFLA Study Group on the Functional Requirements for Bibliographic Records, Saur, München.

Ingwersen, P. and Järvelin, K. (2005), *The Turn: Integration of Information Seeking and Retrieval in Context*, Springer, Dordrecht.

Internet Movie Database (n.d.), available at: www.imdb.com/ (accessed 1 September 2009).

Internet2 Middleware Architecture Committee for Education, Directory Working Group (MACE-Dir) (2008), "EduPerson object class specification", available at: http://middleware.internet2.edu/eduperson/docs/internet2-mace-dir-eduperson-200806.html (accessed 1 September 2009).

ISO 2788 (1986), *Documentation – Guidelines for the Establishment and Development of Monolingual Thesauri*.

ISO 8601 (2004), *Data Elements and Interchange Formats – Information Interchange – Representation of Dates and Times*, 3rd ed.

ISO 14721 (2003), *Space Data and Information Transfer Systems – Open Archival Information System – Reference Model*.

ISO 19107 (2003), *Geographic Information – Spatial Schema*.

ISO 19108 (2002), *Geographic Information – Temporal Schema*.

ISO 21127 (2006), *Information and Documentation – A Reference Ontology for the Interchange of Cultural Heritage Information*.

ISO/IEC 9594-2 (1998), *Information Technology – Open Systems Interconnection – The Directory: Models*, 3rd ed.

ISO/IEC 9594-6 (2005), *Information Technology – Open Systems Interconnection – The Directory: Selected Attribute Types*, 5th ed.

ISO/IEC 15944-1 (2002), *Information Technology – Business Agreement Semantic Descriptive Techniques – Part 1: Operational Aspects of Open-EDI for Implementation*.

ISO/IEC 21000-2 (2005), *Information Technology – Multimedia Framework (MPEG-21) – Part 2: Digital Item Declaration*, 2nd ed.

J. Paul Getty Trust (n.d.), "Getty thesaurus of geographic names", Los Angeles, CA, available at: www.getty.edu/research/conducting_research/vocabularies/tgn/ (accessed 1 September 2009).

James, W. (1890), *The Principles of Psychology*, H. Holt and Company, New York, NY.

Kaptelinin, V. and Nardi, B.A. (2006), *Acting with Technology: Activity Theory and Interaction Design*, MIT Press, Cambridge, MA.

Karasti, H., Baker, K.S. and Halkola, E. (2006), "Enriching the notion of data curation in e-science: data managing and information infrastructuring in the long-term ecological research (LTER) Network", *Computer Supported Cooperative Work*, Vol. 15 No. 4, pp. 321-58.

Kirsh, D. (2001), "The context of work", *Human-Computer Interaction*, Vol. 16 Nos 2-4, pp. 305-22.

Kramer, J., Noronha, S. and Vergo, J. (2000), "A user-centered design approach to personalization", *Communications of the ACM*, Vol. 43 No. 8, pp. 44-8.

Lagoze, C. and Hunter, J. (2001), "The ABC ontology and model", *Journal of Digital Information*, Vol. 2 No. 2.

Lagoze, C., Krafft, D.B., Payette, S. and Jesuroga, S. (2005), "What is a digital library anymore anyway? Beyond search and access in the NSDL", *D-Lib Magazine*, Vol. 11 No. 11.

Lakoff, G. (2008), *The Political Mind: Why You Can't Understand 21st-Century Politics with an 18th-Century Brain*, Viking, New York, NY.

Lakoff, G. and Johnson, M. (1999), *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*, Basic Books, New York, NY.

Latour, B. (1992), "Where are the missing masses? The sociology of a few mundane artifacts", in Bijker, W. and Law, J. (Eds), *Shaping Technology/Building Society: Studies in Sociotechnical Change*, MIT Press, Cambridge, MA, pp. 225-58.

Lea, M., O'Shea, T. and Fung, P. (1995), "Constructing the networked organization: content and context in the development of electronic communications", *Organization Science*, Vol. 6 No. 4, pp. 462-78.

Lee, C.A. (2007), "Taking context seriously: a framework for contextual information in digital collections", UNC SILS TR-2007-04, School of Information and Library Science, University of North Carolina, 18 October, available at: http://sils.unc.edu/research/publications/reports/TR_2007_04.pdf (accessed 1 September 2009).

Lee, C.A., Tibbo, H.R. and Schaefer, J.C. (2007), "Defining what digital curators do and what they need to know: the DigCCurr project", *Proceedings of the 2007 Conference on Digital Libraries*, ACM Press, New York, NY, pp. 49-50.

Léger, L., Tijus, C. and Baccino, T. (2005), "Effect of the task, visual and semantic context on word target detection", in Dey, A., Kokinov, B., Leake, D. and Turner, R. (Eds), *Modeling and Using Context: 5th International and Interdisciplinary Conference CONTEXT 2005, Paris, France, July 5-8, 2005, Proceedings*, Springer, Berlin, pp. 278-91.

Lessig, L. (2006), *Code: Version 2.0*, Basic Books, New York, NY.

Levy, D.M. (2001), *Scrolling Forward: Making Sense of Documents in the Digital Age*, Arcade, New York, NY.

Lexicon Working Group (2000), *GENTECH Genealogical Data Model*, National Genealogical Society, Arlington, VA.

Library and Archives Canada (n.d.), "Business activity structure classification system (BASCS) guidance", Library and Archives Canada, available at: www.collectionscanada.gc.ca/government/products-services/007002-2089-e.html (accessed 1 September 2009).

Liebowitz, S.J. and Margolis, S.E. (2000), "Path dependence", in Bouckaert, B. and Geest, G.D. (Eds), *Encyclopedia of Law and Economics*, Edward Elgar, Cheltenham, pp. 981-98.

Light, M. (2007), "Moving beyond the name: defining corporate entities to support provenance-based access", *Journal of Archival Organization*, Vol. 5 Nos 1/2, pp. 49-74.

Light, M. and Hyry, T. (2002), "Colophons and annotations: new directions for the finding aid", *American Archivist*, Vol. 65 No. 2, pp. 216-30.

Lindblom, J. and Ziemke, T. (2003), "Social situatedness of natural and artificial intelligence: Vygotsky and beyond", *Adaptive Behavior*, Vol. 11 No. 2, pp. 79-96.

Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project (2002), Vancouver, Canada, available at: www.interpares.org/book/index.cfm (accessed 1 September 2009).

McCarthy, J. (1987), "Generality in artificial intelligence", *Communications of the ACM*, Vol. 30 No. 12, pp. 1030-5.

McCarthy, J. (1993), "Notes on formalizing contexts", in Bajcsy, R. (Ed.), *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, Morgan Kaufmann, San Mateo, CA, pp. 555-60.

McDonald, J. (1997), "Towards automated record keeping, interfaces for the capture of records of business processes", *Archives and Museum Informatics*, Vol. 11, pp. 277-85.

McKay, A. and Yakel, E. (2006), "The archival metrics project and beyond: creating tools to share knowledge of archival use and users", paper presented at the ICA-SUV Seminar, 13-20 September, Reykjavik, available at: www2.hi.is/Apps/WebObjects/HI.woa/swdocument/1010097/Aprille_Yakel.pdf (accessed 30 August 2009).

Malone, T.W. (1983), "How do people organize their desks? Implications for the design of office information systems", *ACM Transactions on Information Systems*, Vol. 1 No. 1, pp. 99-112.

Mani, A. and Sundaram, H. (2007), "Modeling user context with applications to media retrieval", *Multimedia Systems*, Vol. 12 Nos 4-5, pp. 339-53.

Matuszek, C., Cabral, J., Witbrock, M. and DeOliveira, J. (2006), "An introduction to the syntax and content of CYC", paper presented at the AAAI 06 Spring Symposium: Formalizing and Compiling Background Knowledge and Its Applications to Knowledge Representation and Question Answering, Stanford, CA, 27-29 March, available at: www.cyc.com/doc/white_papers/AAAI06SS-SyntaxAndContentOfCyc.pdf (accessed 1 September 2009).

Meltzer, S. (2009), *EventsML-G2: Specification Version 1.1*, Core Conformance Level, International Press Telecommunications Council, London.

Merton, R.K. (1957), "The role-set: problems in sociological theory", *British Journal of Sociology*, Vol. 8 No. 2, pp. 106-20.

Miller, C.R. (1994), "Genre as social action", in Freedman, A. and Medway, P. (Eds), *Genre and the New Rhetoric*, Taylor & Francis, Bristol, PA, pp. 23-42.

Miller, D.P. (2000), "Out from under: form/genre access in LCSH", *Cataloging and Classification Quarterly*, Vol. 29 Nos 1/2, pp. 169-88.

Minsky, M. (1974), "A framework for representing knowledge", MIT-AI Laboratory Memo 306, available at: http://hdl.handle.net/1721.1/6089 (accessed 5 March 2010).

Minsky, M.L. (1986), *The Society of Mind*, Simon & Schuster, New York, NY.

Mourelatos, A.P.D. (1978), "Events, processes, and states", *Linguistics and Philosophy*, Vol. 2 No. 3, pp. 415-34.

Naphade, M., Smith, J.R., Tesic, J., Chang, S.-F., Hsu, W., Kennedy, L., Hauptmann, A. and Curtis, J. (2006), "Large-scale concept ontology for multimedia", *IEEE Multimedia*, Vol. 13 No. 3, pp. 86-91.

National Imagery and Mapping Agency (2000), *Department of Defense World Geodetic System 1984: Its Definition and Relationships with Local Geodetic Systems*, 3rd ed., Amendment 1, NIMA Technical Report TR8350.2, National Imagery and Mapping Agency, St Louis, MO.

National Information Standards Organization (2006), *Data Dictionary – Technical Metadata for Digital Still Images*, National Information Standards Organization, Bethesda, MD, (ANSI/NISO Z39.87. 2006).

National Library of Australia (n.d.), "People Australia", National Library of Australia, available at: www.nla.gov.au/initiatives/peopleaustralia/ (accessed August 29, 2009).

National Library of New Zealand (2003), "Metadata standards framework – preservation metadata (revised)", available at: www.natlib.govt.nz/files/4initiatives_metaschema_revised.pdf (accessed 30 August 2009).

National O*NET Consortium (2009), "O*NET-SOC Taxonomy", National O*NET Consortium, available at: www.onetcenter.org/taxonomy.html (accessed 1 September 2009).

Niles, I. and Pease, A. (2001), "Towards a standard upper ontology", in Welty, C. and Smith, B. (Eds), *Formal Ontology in Information Systems: Collected Papers from the 2nd International Conference, the Cliff House on Bald Head Cliff Overlooking the Atlantic Ocean, October 17-19, 2001*, ACM Press, New York, NY, pp. 2-9.

Nordland, L. (2004), "The concept of 'secondary provenance': re-interpreting Ac Ko Mok Ki's Map as evolving text", *Archivaria*, Vol. 58, pp. 147-59.

Norman, D.A. (1990), *The Design of Everyday Things*, Doubleday, New York, NY.

North American Industry Classification System (2007), US Census Bureau, Washington, DC, available at: www.census.gov/eos/www/naics/ (accessed 1 September 2009).

O'Donnell, C.R., Tharp, R.G. and Wilson, K. (1993), "Activity settings as the unit of analysis: a theoretical basis for community intervention and development", *American Journal of Community Psychology*, Vol. 21 No. 4, pp. 501-20.

Orlikowski, W.J. (2000), "Using technology and constituting structures: a practice lens for studying technology in organizations", *Organization Science*, Vol. 11 No. 4, pp. 404-28.

Oxford Dictionary of National Biography (n.d.), Oxford University Press, available at: www.oxforddnb.com/ (accessed 1 September 2009).

Oxford English Dictionary (1989), 2nd ed., Oxford University Press, Oxford.

Palmer, C., Zavalina, O. and Mustafoff, M. (2007), "Trends in metadata practices: a longitudinal study of collection federation", in Larson, R., Rasmussen, E., Sugimoto, S. and Toms, E. (Eds), *Proceedings of the 7th ACM/IEEE Joint Conference on Digital Libraries, Vancouver, British Columbia, Canada, June 18-23, 2007*, ACM Press, New York, NY, pp. 386-95.

Panizzi, A. (Ed.) (1841), *Catalog of Printed Books in the British Museum*, Vol. 1, J.B. Nichols and Son, London.

Patton, G.E. (Ed.) (2009), *Functional Requirements for Authority Data: A Conceptual Model*, K.G. Saur, München.

Pearce-Moses, R. (2005), *Context, in A Glossary of Archival and Records Terminology*, Society of American Archivists, Chicago, IL, pp. 90-1.

Peirce, C.S. (1880), "On the algebra of logic", *American Journal of Mathematics*, Vol. 3 No. 1, pp. 15-57.

Penker, M. and Eriksson, H.-E. (2000), *Business Modeling with UML: Business Patterns at Work*, John Wiley & Sons, New York, NY.

Petras, V., Larson, R.R. and Buckland, M. (2006), "Time period directories: a metadata infrastructure for placing events in temporal and geographic context", *Opening Information Horizons: 6th ACM/IEEE-CS Joint Conference on Digital Libraries: June 11-15, 2006, Chapel Hill, NC, USA: JCDL 2006*, ACM Press, New York, NY, pp. 151-60.

Powell, A., Heaney, M. and Dempsey, L. (2000), "RSLP collection description", *D-Lib Magazine*, Vol. 6 No. 9.

PREMIS Working Group (2005), "Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group", OCLC Online Computer Library Center and Research Libraries Group, available at: www.oclc.org/research/projects/pmwg/premis-final.pdf (accessed 1 September 2009).

Pustejovsky, J., Castaño, J., Ingria, R., Saurí, R., Gaizauskas, R., Setzer, R. and Katz, G. (2003), "TimeML: robust specification of event and temporal expressions in text", in Bunt, H., van der Sluis, I. and Morante, R. (Eds), *Proceedings of the IWCS-5 5th International Workshop on Computational Semantics, Tilburg University, Computational Linguistics and AI Group, Tilburg, The Netherlands*.

Quine, W.V. (1951), *Mathematical Logic*, Rev. ed., Harvard University Press, Cambridge, MA.

Rajasekar, A., Wan, M., Moore, R., Schroeder, W., Chen, S-Y., Gilbert, L., Hou, C-Y., Lee, C.A., Marciano, R., Tooby, P., de Torcy, A. and Zhu, B. (2010), *iRODS Primer: Integrated Rule-Oriented Data System*, Morgan & Claypool, San Rafael, CA.

Ranganathan, S.R. (1938), *Theory of Library Catalogue*, Madras Library Association, Madras.

Ranganathan, S.R. (1965), *The Colon Classification*, Vol. 4, Graduate School of Library Service Rutgers the State University, New Brunswick, NJ.

Ranganathan, S.R. and Gopinath, M.A. (1967), *Prolegomena to Library Classification*, 3rd ed., Asia Publishing House, London.

Rheingold, H. (2003), *Smart Mobs: The Next Social Revolution*, Perseus, Cambridge, MA.

Ringel, M., Cutrell, E., Dumais, S. and Horvitz, E. (2003), "Milestones in time: the value of landmarks in retrieving information from personal stores", in Rauterberg, M., Menozzi, M. and Wesson, J. (Eds), *Human-Computer Interaction, INTERACT '03: IFIP TC13 International Conference on Human-Computer Interaction, Zurich, Switzerland, 1-5 September*, IOS Press, Amsterdam, pp. 184-91.

Robertson, B. (2009), "Exploring historical RDF with HEML", *Digital Humanities Quarterly*, Vol. 3 No. 1.

Robertson, D.W. Jr (1946), "A note on the classical origin of 'circumstances' in the *Medieval Confessional*", *Studies in Philology*, Vol. 43 No. 1, pp. 6-14.

Robinson, C. (1997), "Records control and disposal using functional analysis", *Archives and Manuscripts*, Vol. 25 No. 2, pp. 288-303.

Roe, K. (1992), "Enhanced authority control: is it time?", *Archivaria*, Vol. 35, pp. 119-29.

Roe, K. (2005), *Arranging & Describing Archives & Manuscripts*, Society of American Archivists, Chicago, IL.

Rogers, E.M. (1995), *Diffusion of Innovations*, 4th ed., Free Press, New York, NY.

Russell, B. (1938), *Principles of Mathematics*, Norton & Company, New York, NY.

Rust, G. and Bide, M. (2000), "The < indecs > metadata framework: principles, model and data dictionary", WP1a-006-2.0, available at: www.doi.org/topics/indecs/indecs_framework_2000.pdf (accessed 30 August 2009).

Saarinen, E. (1956), "The maturing modern", *Time*, 2 July, p. 54.

Schellenberg, T.R. (1956), *Modern Archives: Principles and Techniques*, University of Chicago Press, Chicago, IL.

Scheufele, D.A. (1999), "Framing as a theory of media effects", *Journal of Communication*, Vol. 49 No. 1, pp. 103-22.

Schilit, B.N. and Theimer, M.M. (1994), "Disseminating active map information to mobile hosts", *IEEE Network*, Vol. 8 No. 5, pp. 22-32.

Scott, P.J. (1966), "The record group concept: a case for abandonment", *American Archivist*, Vol. 29, pp. 493-504.

Scriberras, A. (Ed.) (2006), "Lightweight directory access protocol (LDAP): schema for user applications" (RFC 4519), memo, Network Working Group.

Sellen, A.J. and Harper, R. (2002), *The Myth of the Paperless Office*, MIT Press, Cambridge, MA.

Shanon, B. (1990), "What is context?", *Journal for the Theory of Social Behaviour*, Vol. 20 No. 2, pp. 157-66.

Sharer, R.J. and Ashmore, W. (1987), *Archaeology: Discovering our Past*, Mayfield, Palo Alto, CA.

Sharp, A. and McDermott, P. (2009), *Workflow Modeling: Tools for Process Improvement and Applications Development*, Artech House, Boston, MA.

Shatford, A. (1986), "Analyzing the subject of a picture: a theoretical approach", *Cataloging & Classification Quarterly*, Vol. 6 No. 3, pp. 39-62.

Smith, B. (2007), "On place and space: the ontology of the ERUV", in Kanzian, C. and Runggaldier, E. (Eds), *Cultures: Conflict – Analysis – Dialogue; Proceedings of the 29th International Ludwig Wittgenstein Symposium, Kirchberg am Wechsel, Austria 2006*, Ontos, Frankfurt, pp. 403-16.

Smith, B. and Varzi, A.C. (2000), "Fiat and bona fide boundaries", *Philosophy and Phenomenological Research*, Vol. 60 No. 2, pp. 401-20.

Smith, B.C. (1996), *On the Origin of Objects*, MIT Press, Cambridge, MA.

Smith, D.A. (2002), "Detecting and browsing events in unstructured text", in Beaulieu, M., Baeza-Yates, R., Myaeng, S.H. and Jarvelin, K. (Eds), *SIGIR 2002: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Tampere, Finland, August 11-15*, ACM Press, New York, NY, pp. 73-80.

Smith, D.A. and Crane, G. (2001), "Disambiguating geographic names in a historical digital library?", in Constantopoulos, P. and Sølvberg, I. (Eds), *Research and Advanced Technology for Digital Libraries 5th European Conference, ECDL 2001, Darmstadt, Germany, September 4-9*, Springer, Berlin, pp. 127-36.

Smythe, C., Tansey, F. and Robson, R. (2001), *IMS Learner Information Packaging Information Model Specification*, Version 1.0, IMS Global Learning Consortium, Lake Mary, FL.

Snow, D.A., Rochford, E.B., Worden, S.K. and Benford, R.D. (1986), "Frame alignment processes, micromobilization, and movement participation", *American Sociological Review*, Vol. 51 No. 4, pp. 464-81.

Snow, D.R., Gahegan, M., Giles, C.L., Hirth, K.G., Milner, G.R., Mitra, P. and Wang, J.Z. (2006), "Cybertools and archeology", *Science*, Vol. 311 No. 5763, pp. 958-9.

Society of American Archivists (2004), *Describing Archives: A Content Standard*, Chicago, IL.

Society of American Archivists and Library of Congress (2002), "Encoded Archival Description Tag Library", available at: www.loc.gov/ead/tglib/ (accessed 1 September 2009).

Souza, C.D., Froehlich, J. and Dourish, P. (2005), "Seeking the source: software source code as a social and technical artifact", in Schmidt, K., Pendergast, M., Ackerman, M. and Mark, G. (Eds), *GROUP '05: Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work, Sanibel Island, FL, USA, November 6-9*, Association for Computing Machinery, New York, NY, pp. 197-206.

Sperberg-McQueen, C.M. and Burnard, L. (Eds) (2002), "TEI P4: guidelines for electronic text encoding and interchange", Text Encoding Initiative Consortium, available at: www.tei-c.org/P4X/ (accessed 1 September 2009).

Stalnaker, R.C. (1978), "Assertion", *Syntax and Semantics*, Vol. 9, pp. 315-22.

Stalnaker, R. (1998), "On the representation of context", *Journal of Logic, Language and Information*, Vol. 7 No. 1, pp. 3-19.

Standard Occupational Classification System (2000), US Department of Labor, Bureau of Labor Statistics, Washington, DC, available at: http://stats.bls.gov/soc/ (accessed 1 September 2009).

Suchman, L.A. (1987), *Plans and Situated Actions: The Problem of Human-Machine Communication*, Cambridge University Press, Cambridge.

Suderman, J. (2001), "Context, structure and content: new criteria for appraising electronic records", *Annual Conference of the Association of Canadian Archivists, Winnipeg, Manitoba, 4-9 June*, available at: www.rbarry.com/suderman-wholepaper7_postscript011102.htm (accessed 1 September 2009).

Suggested Upper Merged Ontology (SUMO) (n.d.), available at: www.ontologyportal.org/ (accessed 30 August 2009).

Sullivan, T., Irvine, A. and Haas, S.W. (2008), "It's all relative: usage of relative temporal expressions in triage notes", *Proceedings of the American Society for Information Science and Technology*, Vol. 45 No. 1.

Sweet, M. and Thomas, D. (2000), "Archives described at collection level", *D-Lib Magazine*, Vol. 6 No. 9.

Szary, R.V. (2006), "Encoded archival context (EAC) and archival description: rationale and background", *Journal of Archival Organization*, Vol. 3 Nos 2/3, pp. 217-27.

Talja, S., Keso, H. and Pietiläinen, T. (1999), "The production of 'context' in information-seeking research: a metatheoretical view", *Information Processing and Management*, Vol. 35 No. 6, pp. 751-63.

Tarski, A. (1941), "On the calculus of relations", *Journal of Symbolic Logic*, Vol. 6 No. 3, pp. 73-89.

Taylor, A.G. and Miller, D.P. (2006), *Introduction to Cataloging and Classification*, 10th ed., Libraries Unlimited, Westport, CT.

Thibodeau, S. (1995), "Archival context as archival authority record: the ISAAR(CPF)", *Archivaria*, Vol. 40, pp. 75-85.

Thomas, W., Gregory, A., Kuo, I.-L., Wackerow, J. and Nelson, C. (Eds) (2008), *Data Documentation Initiative (DDI) Technical Specifications, Part I: Overview (Version 3.0)*, DDI Alliance.

Tibbo, H.R. (1993), *Abstracting, Information Retrieval, and the Humanities*, American Library Association, Chicago, IL.

Transportation Research Board (2008), *Transportation Research Thesaurus*, Transportation Research Board, available at: http://ntlsearch.bts.gov/tris/trt.do (accessed 30 August 2009).

Tyler, A. and Evans, V. (2003), *The Semantics of English Prepositions: Spatial Scenes, Embodied Meaning, and Cognition*, Cambridge University Press, Cambridge.

US Board on Geographic Names (n.d.), "Geographic names information system", US Board on Geographic Names, available at: http://geonames.usgs.gov/ (accessed 1 September 2009).

US Department of Education (n.d.), "ERIC thesaurus", US Department of Education, Institute of Education Sciences, available at: www.eric.ed.gov/thesaurus (accessed 1 September 2009).

US Library of Congress (n.d.a), "Metadata Authority Description Schema (MADS)", Washington, DC, available at: www.loc.gov/standards/mads/ (accessed 1 September 2009).

US Library of Congress (n.d.b), "Metadata Object Description Schema (MODS)", US Library of Congress, Washington, DC, available at: www.loc.gov/standards/mods/ (accessed 1 September 2009).

US Library of Congress (n.d.c), "Name authority cooperative program of the program for cooperative cataloging", US Library of Congress, Washington, DC, available at: www.loc.gov/catdir/pcc/naco/ (accessed 1 September 2009).

US Library of Congress (n.d.d), "Source codes for genre", Library of Congress, Washington, DC, available at: www.loc.gov/marc/sourcecode/genre/genresource.html (accessed 1 September 2009).

US Library of Congress (2003), "MARC value list for relators and roles", Washington, DC, available at: www.loc.gov/marc/sourcecode/relator/relatorlist.html (accessed 1 September 2009).

US Library of Congress (2010), "MARC 21 format for bibliographic data", US Library of Congress, Washington, DC, available at: www.loc.gov/marc/bibliographic/ecbdhome.html (accessed 20 March 2010).

van Dijk, T.A. (1981), "Episodes as units of discourse analysis", in Tannen, D. (Ed.), *Analyzing Discourse: Text and Talk*, Georgetown University Press, Washington, DC, pp. 177-95.

Vickery, B.C. (1960), *Faceted Classification: A Guide to Construction and Use of Special Schemes*, Aslib, London.

Visual Resources Association (2007), *VRA Core 4.0*, Visual Resources Association, available at: www.vraweb.org/projects/vracore4/index.html (accessed 1 September 2009).

Vitali, S. (2006), "What are the boundaries of archival context? The SIASFI Project and the Online Guide to the Florence State Archives, Italy", *Journal of Archival Organization*, Vol. 3 No. 2, pp. 243-60.

Wallace, D.A. (1995), "Managing the present: metadata as archival description", *Archivaria*, Vol. 39, pp. 22-32.

Wang, X., Demartini, T., Wragg, B., Paramasivam, M. and Barlas, C. (2005), "The MPEG-21 rights expression language and rights data dictionary", *IEEE Transactions on Multimedia*, Vol. 7 No. 3, pp. 408-17.

Wasserman, S. and Faust, K. (1994), *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge.

Weber, M.B. and Favaro, S. (2007), "Beyond Dublin Core: development of the workflow management system and metadata implementation at Rutgers, The State University of New Jersey", *DigCCurr2007: An International Symposium in Digital Curation, Chapel Hill, NC, 18-20 April*, available at: www.ils.unc.edu/digccurr2007/papers/weberFavaro_paper_4-1.pdf (accessed 1 September 2009).

Wedervang-Jensen, E. and Driscoll, M. (2006), "Report on XML mark-up of biographical and prosopographical data", Text Encoding Initiative, available at: www.tei-c.org/Activities/Workgroups/PERS/persw02.xml (accessed 1 September 2009).

Weick, K.E. (1995), *Sensemaking in Organizations*, Sage, Thousand Oaks, CA.

White, S.A. (2008), *Business Process Modeling Notation, v.1.1*, Object Management Group, Needham, MA.

Whitehead, A.N. and Russell, B. (1927), *Principia Mathematica*, 2nd ed., Cambridge University Press, Cambridge.

Wiberley, S. and Jones, W.G. (1989), "Patterns of information seeking in the humanities", *College & Research Libraries*, Vol. 50 No. 6, pp. 638-45.

Wikipedia (2009), "Glossary of systems theory", available at: http://en.wikipedia.org/wiki/Glossary_of_systems_theory (accessed 1 September 2009).

Wilson, A. and Clayphan, R. (2004), "Functional requirements for describing agents", Draft 2, Dublin Core Metadata Initiative – Agents Working Group, available at: http://dublincore.org/groups/agents/agentFRdraft2-2.html (accessed 30 August 2009).

Winner, L. (1986), "Do artifacts have politics?", in Winner, L. (Ed.), *The Whale and the Reactor: A Search for Limits in an Age of High Technology*, University of Chicago Press, Chicago, IL, pp. 19-39.

Winograd, T. and Flores, F. (1986), *Understanding Computers and Cognition: A New Foundation for Design*, Ablex, Norwood, NJ.

Workflow Management Coalition (2008), *XML Process Definition Language*, Version 2.1, Workflow Management Coalition, Hingham, MA.

WorldCat Identities (2008), OCLC Online Computer Library Center, Dublin, OH, available at: www.worldcat.org/identities/ (accessed 1 September 2009).

Yang, M. and Marchionini, G. (2005), in Kellogg, W.A., Zhai, A., Gale, C. and van der Veer, G.C. (Eds), *CHI 2005: Technology, Safety, Community: Conference Proceedings: Conference on Human Factors in Computing Systems: Portland, OR, USA, April 2-7*, Association for Computing Machinery, New York, NY, pp. 1877-80.

Yates, J. and Orlikowski, W. (1992), "Genres of organizational communication: a structurational approach to studying communication and media", *Academy of Management Review*, Vol. 17 No. 2, pp. 299-326.

Yi, K., Beheshti, J., Cole, C., Leide, J.E. and Large, A. (2006), "User search behavior of domain-specific information retrieval systems: an analysis of the query logs from PsycINFO and ABC-Clio's Historical Abstracts/America: history and life: research articles", *Journal of the American Society for Information Science and Technology*, Vol. 57 No. 9, pp. 1208-20.

Zeilenga, K.D. (Ed.) (2006), *COSINE LDAP/X.500 Schema (RFC 4524)*, memo, Network Working Group.

Zimmerman, A.S. (2008), "New knowledge from old data: the role of standards in the sharing and reuse of ecological data", *Science, Technology, & Human Values*, Vol. 33 No. 5, pp. 631-52.

Further reading

Consultative Committee for Space Data Systems (1992), *Report Concerning Space Data System Standards: Standard Formatted Data Units – A Tutorial, Green Book*, Consultative Committee for Space Data Systems, Washington, DC, Issue 1.

ISO/CD 27729 (2008), Information and Documentation – International Standard Name Identifier (ISNI), National Information Standards Organization (NISO), Baltimore, MD.

National O*NET Consortium (n.d.), "O*NET Content Model", National O*NET Consortium, available at: www.onetcenter.org/content.html (accessed 1 September 2009).

Preyer, G. and Peter, G. (2005), *Contextualism in Philosophy: Knowledge, Meaning, and Truth*, Oxford University Press, New York, NY.

Schilit, B.N., Adams, N. and Want, R. (1995), "Context-aware computing applications", in Cabrera, L-F. and Satyanarayanan, M. (Eds), *Workshop on Mobile Computing Systems and Applications: Proceedings, Santa Cruz, CA, December 8-9, 1994*, IEEE Computer Society, Los Alamitos, CA, pp. 85-90.

Appendix. Guidance and considerations for generating or capturing information about the nine classes of contextual entities

The framework in this paper draws from a diverse range of sources. There is a general overview of the sources in the section of the paper called "Building blocks and motivation for the contextual information framework." This appendix provides considerably more detail about the specific sources, with a particular emphasis on those that articulate specific descriptive conventions and practices.

International Standard Archival Authority Record for Corporate Bodies, Persons and Families (ISAAR/CPF) (ICA Committee on Descriptive Standards, 2004) and Encoded Archival Context – Corporate bodies, Persons, and Families (EAC-CPF) are two recent efforts specifically to formalize contextual information related to archival materials (http://eac.staatsbibliothek-berlin.de/) (Szary, 2006). The Research Support Libraries Programme (RSLP) has also produced several guidance documents for collection-level description (Powell *et al.*, 2000).

Several high-level conceptual models or ontologies provide useful building blocks for a contextual information framework. These include the ABC Ontology (Lagoze and Hunter, 2001); International Committee for Documentation Conceptual Reference Model (CIDOC CRM) (ISO 21127, 2006); Cyc (Guha and Lenat, 1994); Suggested Upper Merged Ontology (SUMO) (Niles and Pease, 2001); Large Scale Concept Ontology for Multimedia (LSCOM) (Naphade *et al.*, 2006); Functional Requirements for Bibliographic Records (FRBR) (IFLA Study Group, 1988); and Functional Requirements for Authority Data (Patton, 2009).

Many descriptive and metadata standards also specify classes, entities, elements, properties and attributes that can serve as contextual information for a target object. Several that have informed this paper are: Australian Government Recordkeeping Metadata Standard (2008); Data Dictionary-Technical Metadata for Digital Still Images (National Information Standards Organization, 2006); Describing Archives: A Content Standard (DACS) (Society of American of American Archivists, 2004); Data Documentation Initiative (DDI) Version 3 (Thomas *et al.*, 2008); Encoded Archival Description (EAD) (Society of American Archivists and Library of Congress, 2002); GENTECH Genealogical Data Model (Lexicon Working Group, 2000); Metadata Encoding and Transmission Standard (Digital Library Federation, 2007); International Standard Archival Authority Record for Corporate Bodies, Persons and Families (ISAAR/CPF) (ICA Committee on Descriptive Standards, 2004); General International Standard Archival Description (ISAD(G)) (Ad Hoc Commission on Descriptive Standards, 2000); Metadata Authority Description Schema (MADS) (US Library of Congress, n.d.a); Metadata Object Description Standard (MODS) (US Library of Congress, n.d.b); MPEG-21 (ISO/IEC 21000-2, 2005; Wang *et al.*, 2005); Metadata Standards Framework – Preservation Metadata (National Library of New Zealand, 2003); PREservation Metadata: Implementation Strategies (PREMIS) (PREMIS Working Group, 2005); RUCore Data Model (Weber and Favaro, 2007); Resource Description and Access (RDA) (American Library Association, 2008); Rules for Archival Description (RAD) (Canadian Council of Archives, 2003); Text Encoding Initiative (TEI) (Sperberg-McQueen and Burnard, 2002); Transportation Research Thesaurus (Transportation Research Board, 2008); PRISM (IDEAlliance, 2006); VRA Core 4.0 (Visual Resources Assocaition, 2007); and Categories for the Description of Works of Art (CDWA) (Baca and Harpring, 2009).

There are several other relevant standards that define and potentially support the interactions between a diverse set of entities over time. The GENTECH Genealogical Data Model (Lexicon Working Group, 2000) supports assertions between four types of entities: persona, event, characteristic, and group. Standards related to digital rights management – Interoperability of Data in E-commerce Systems (INDECS) and MPEG-21 Rights Expression Language – provide fruitful building blocks for a model of contextual information (Rust and Bide, 2000; Wang *et al.*, 2005). The Lightweight Directory Access Protocol (LDAP) based on the × .500 family, is a well-established set of standards for identifying and locating "anything which is identifiable (can be named)" (ISO/IEC 9594-2, 1998). LDAP provides the following classes of digital objects (among others): application process, country, device, locality, organization,

organizational person, organizational role, organizational unit, and person (Scriberras, 2006; Zeilenga, 2006).

The Perseus Project has provided substantial innovation in identifying and exploiting information related to contextual entities. The project's services allow for searching on place, person, date or date range, and it has investigated named entity recognition for a variety of entity types: areas, currency, dates, events, geographic names (places), measures, newspapers, organizations, personal names, products, railroads, regiments, ships, street names and addresses (Crane and Jones, 2006; Smith, 2002; Smith and Crane, 2001).

Counting Online Usage of Networked Electronic Resources (COUNTER, n.d.) and the Developing Archival Metrics Project (McKay and Yakel, 2006) both provide useful steps toward more systematic capture, identification and management of use data. Measuring the Impact of Networked Electronic Services (MINES) is a methodology for making inferences about the purposes and demographics of electronic resource users based on survey data collected at points of use (Franklin and Plum, 2006).

The following is a discussion of considerations and available sources of guidance for representing contextual items associated with each of the nine classes of entities in the contextual information framework.

1.1 Object

This category of contextual entities includes both digital and physical objects. There are numerous sources of guidance for representing information about physical objects, ranging from the Global Trade Item (GTIN) for commercial products (GS1 US, 2006), to the Categories for the Description of Works of Art (CDWA) for art and material culture, to the relatively institution-specific conventions for representing archeological artifacts (Snow *et al.*, 2006). No repository will fully adopt all of these standards. However, in cases when collections of target digital objects (TDOs) could be significantly enhanced by providing contextual information associated with physical objects (i.e. when many of the important contextual entities are physical objects), curators of those collections could benefit from the standards, in order to either directly incorporate or link to contextual items related to the physical objects.

As information professionals have taken on responsibility for digital objects of increasing complexity, they have developed and adopted numerous conventions for representing and documenting that complexity. Most of these conventions were designed for specific domains or object types, e.g. Data Documentation Initiative (DDI) for social science data; Standard Formatted Data Unit (SFDU) (Consultative Committee for Space Data Systems, 1992), and its intended successor XML Formatted Data Unit (XFDU), for space and terrestrial data; MPEG-21 for video; and IMS Learner Information Package Information Model Specification (Smythe *et al.*, 2001) for learning objects. However, the specific origin of specifications does not preclude their use in other domains or for other types of objects. In fact, advocates of all the above specifications have proposed that they can be used, or at least serve as models, in other areas (see, e.g. Bekaert *et al.*, 2005). The Metadata Encoding and Transmission Standard (METS) (Digital Library Federation, 2007), on the other hand, is an example of a specification that has been designed from its inception to have a very broad scope of application.

The development of conventions for representing object-level complexity has occurred at the same time that information professionals have been moving away from intensive item-level (or even series-level or sub-collection-level) description, because of the massive volume of materials, limited institutional resources, and many new service expectations that draw on those resources. In order to address this apparent contradiction in practices and priorities, curators of digital collections will need to adopt innovative automated (or semi-automated) and aggregate-level approaches for representing and documenting object-level characteristics. Two essential sources of contextual information for a digital object are its external relations to other objects (see relationship below) and its "internal compositional," which is the way the components that make up the object are arranged and associated with each other (Hedstrom and Lee, 2002).

1.2 Agent

Descriptive standards have traditionally focused more on information resources – such as bibliographic units – than the agents who interact with them. Information about agents has, therefore, often been embedded within bibliographic utilities and standards, rather than being conceptualized separately. However, librarians and archivists have been working for some time on the elusive goal of uniquely identifying and describing agents over time. An Agents Working Group was formed in 1998, in order to address the agent information that was potentially embedded in (or missing from) the Dublin Core elements (Wilson and Clayphan, 2004). A project within the International Organization for Standardization (ISO) is developing the International Standard Name Identifier (ISNI) (ISO/CD 27729) to uniquely identify "public identities" across multiple areas of creative activity. International Standard Archival Authority Record for Corporate Bodies, Persons, and Families (ISAAR(CPF) (ICA Committee on Descriptive Standards, 2004) and Encoded Archival Context – Corporate bodies, Persons, and Families (EAC-CPF) (http://eac.staatsbibliothek-berlin.de/) (Szary, 2006) are two rich sources of guidance on the types of information one might hope to provide about agents. RDA (American Library Association, 2008) also provides detailed guidance for recording attributes of persons, families, and corporate bodies. In 2006, the Text Encoding Initiative also initiated the Personography Task Force, one product of which has been a report that describes and compares many existing schemes for marking up information about individuals (Wedervang-Jensen and Driscoll, 2006).

When creating authority records and other metadata associated with "corporate bodies" as agents, a vital issue is how to "draw boundaries around one entity and distinguish it from others," particularly when the names, functions and internal structures change over time. The guidance for identifying and representing such boundaries varies substantially across different descriptive standards (Light, 2007).

The Library of Congress has maintained the Name Authority File for this purpose, accepts name authority records from other institutions through its Name Authority Cooperative (NACO) program (US Library of Congress, n.d.c), and has joined with other major libraries from the English-speaking world to develop the Anglo-American Authority File (AAAF). The Deutsche Nationalbibliothek, US Library of Congress, and Bibliothèque Nationale de France are working with OCLC to provide unified access to their authority records through the Virtual International Authority File (http://viaf.org/). ONE Shared Authority Control (ONESAC) "consolidates bibliographic authority data about authors, corporations, conferences, subjects and thesauri" of European libraries (www.portia.dk/websites/onesac.htm). The National Library of Australia (n.d.) is developing a resource called People Australia, which "will allow users to access information about significant Australian people and organisations as well as related biographical and contextual information." MARC 21 (US Library of Congress, 2010) provides fields for Personal Names (100, 600, 700), Corporate Names (110, 610, 710, 810), Participants or Performers (511), Biographical or Historical Data (545), and Issuing Body (550). Valuable source of biographical and name information include the Biographical Directory of the United States Congress (n.d.), Internet Movie Database (n.d.), Oxford Dictionary of National Biography (n.d.), Union List of Artist Names (ULAN) (Harpring *et al.*, 2006), and WorldCat Identities (WorldCat Identities, 2008).

There are many other efforts to specify information about agents to support interchange, discovery and reuse across the internet. vCard, for example, is a directory profile for the representation and exchange of information about individuals, including identification and naming; addressing; geographical positions or regions; and place or role within an organization (Dawson and Howes, 1998), which can be embedded into Extensible Hypertext Markup Language (XHTML) documents using hCard. X.520, X.521 (ISO/IEC 9594-6, 2005; ISO/IEC 9594-7), LDAP (Scriberras, 2006; Zeilenga, 2006) and EduPerson (MACE-Dir, 2008) specify a number of element and attribute types for describing agents. The Friend of a Friend (FOAF) vocabulary defines a set of classes and properties for encoding information on web pages about individuals and associated entities, such as documents, groups, online accounts, organizations,

projects (Brickley and Miller, 2007). OpenID defines "eight commonly requested pieces of information" about individuals (Hoyt *et al.*, 2006). Another source of guidance is the work on user modeling, including the General User Model Ontology (GUMO) (Heckmann *et al.*, 2005), the User Modeling Markup Language (UserML) (Heckmann and Krüger, 2003), and IMS Learner Information Package Information Model Specification (Smythe *et al.*, 2001).

There are numerous ways to classify agents. Some of the most influential metadata schemes for digital collections identify types of agents, but leave the typology quite simple. METS allows for individual, organization, or other. PREservation Metadata: Implementation Strategies (PREMIS) (PREMIS Working Group, 2005) suggests: person, organization or software. The Library of Congress has also developed the Metadata Authority Description Schema (MADS), which is an XML schema for an authority element set for "metadata about agents (people, organizations), events, and terms (topics, geographics, genres, etc)." O*NET supports detailed specification of jobs and responsibilities through its Content Model and taxonomy, the latter being based on the Standard Occupational Classification (SOC) System of the US Bureau of Labor Statistics. The ERIC Thesaurus (US Department of Education, n.d.) also includes a category devoted to "occupations."

1.3 Occurrence

There is a growing body of building blocks for the identification and encoding of occurrence information. Guidance for the detailed representation of processes includes the Process Specification Language (Bock and Gruninger, 2005); extension and application of the Unified Modeling Language (Penker and Eriksson, 2000); XML Process Definition Language (Workflow Management Coalition, 2008); and the Business Process Modeling Notation Specification (White, 2008). TimeML and the Historical Event Markup and Linking (HEML) Project provide conventions for encoding and storage of event information. TimeML is designed to support time stamping and ordering of events, as well as reasoning about "contextually underspecified temporal expressions" and the persistence of events (Pustejovsky *et al.*, 2003); and the HEML schema includes elements for location, time, persons, roles and evidence for the event (Robertson, 2009). PRISM defines metadata fields "to provide information about an event pictured in the image or contributing to the image" (IDEAlliance, 2006). One of the ultimate goals of MPEG-21 is to address "event reporting." The MPEG-21 Rights Data Dictionary (RDD) elaborates 14 types of Acts that can be performed on resources (Wang *et al.*, 2005). ISO/IEC 15944-1 (2002) presents an approach for representing occurrences, particularly formal business transactions. The Union List of Artist Names (ULAN) provides detailed guidance for documenting "a critical event, activity, state or status, or situation in the person's life or the corporate body's history." (Harpring *et al.*, 2006). EventsML-G2 is a detailed specification for describing and sharing information about events, designed for event planners and news reporters (Meltzer, 2009). The Internet Calendaring and Scheduling Core Object Specification (iCalendar) identifies numerous components and parameters associated with single or recurring events (Dawson and Stenerson, 1998). CIDOC CRM and the GENTECH Genealogical Data Model also specify ways to represent events. MARC 21 (US Library of Congress, 2010) includes fields (033 and 518) for Date/Time and Place of an Event associated with "creation, capture, recording, filming, execution, or broadcast associated with an event or the finding of a naturally occurring object"; fields (611, 711) for documenting a meeting as a type of event; an Action Note field (583) for "processing, reference, and preservation actions"; and subfield codes for location of an event ($c for field 650) and for the security downgrading or declassification of an item ($d for field 355) as a particular type of event.

1.4 Purpose

Functions often have hierarchical or nested relationships with other functions. Two sources of guidance for representing functional entities and their relationships from Australia are the Australian Governments' Interactive Functions Thesaurus (2007) and Keyword AAA (Robinson, 1997), and one from Canada is the Business Activity Structure Classification System (BASCS)

Guidance (Library and Archives Canada, n.d.). The International Standard for Describing Functions (ICA Committee on Best Practices and Standards, 2008) has been designed to describe functions within archival information systems. The Australian Government Recordkeeping Metadata Standard also includes two relevant entities: Business, which is "a business function, activity or transaction performed by, or assigned to, an organisation or its employees", and Mandate, which is "a source of business requirements, including recordkeeping requirements." There are numerous sets of conventions for representing purposes, functions and mandates within specific governmental, institutional or organizational contexts (e.g. codes of regulations, policies, budget codes, procedures manuals, strategic planning documents), all of which can serve as rich sources of contextual information associated with digital objects. MARC 21 (US Library of Congress, 2010) includes fields for Function (657) and Funding Information (536).

1.5 Time

The most straightforward case of representing time is a precise time and date, as specified in ISO 8601 (2004). However, there is a myriad of other possible temporal units and expressions, which TIMEX2 attempts to accommodate (Ferro *et al.*, 2005). ISO 19108 (2002) provides detailed guidance for representing "temporal feature attributes, feature operations, and feature associations, and for defining the temporal aspects of metadata about geographic information," though it is potentially applicable for describing other types of information. The Time Period Directory initiative aims to support translations between common language labels, such as the Civil War, and specific time spans (Petras *et al.*, 2006). There are many other relevant specifications and research activities that fall within the arena of "temporal modeling," which attempt to address the deep connections between events (see above) and time (e.g. Grandi *et al.*, 2005). In order to convey contextual information associated with a digital object, it can often be valuable to reflect an intersection between time and one or more of the other contextual entities.

1.6 Place

There are a number of detailed standards and guidance documents for encoding place information. The Alexandria Digital Library project offers a "Guide to the ADL Gazetteer Content Standard" (Alexandria Digital Library Project, 2004). A well-established set of conventions for encoding locations as coordinates is available in the Department of Defense World Geodetic System 1984 (National Imagery and Mapping Agency, 2000), which is supported by vCard and the geo microformat (Çeelik, 2007). vCard also allows for specifying location based on time zone. The X.500 and LDAP families of standards identify ways to encode geographic and postal addresses. There are several detailed elaborations of places and types of places, including the Alexandria Digital Library Feature Type Thesaurus (2002), Geographic Names Information System (US Board on Geographic Names, n.d.), and the Getty Thesaurus of Geographic Names (TGN) (J. Paul Getty Trust, n.d.). The creation of geo-referenced data is increasing dramatically, not only through direct entry into dedicated geographic information systems (GIS), but also through automatic capture by devices such as digital cameras or later assigned by users as "geotags." As Howard Rheingold points out, "Knowing our exact geographic location is one form of context awareness in which machines are better than humans" (Rheingold, 2003, p. 97). Conventions for identifying place will be important for supporting the interoperability and reuse of the place data. Although it is closely connected to specific geographic localities, "nationality" (Harpring *et al.*, 2006) is usually best considered a characteristic of an agent, because it is more a statement of personal identity and status than a clear indication of where someone was born, was raised, or currently lives. MARC 21 (US Library of Congress, 2010) provides fields for Geographic Name (651, 751) and Hierarchical Place (752) added entries and Geographic Coverage notes (522).

1.7 Form of expression

There can be value in distinguishing between information about purposes and form of expression associated with a TDO, though in traditional archival descriptive practice, these two types of contextual entities have often been combined in ways that can be difficult to disentangle (Bearman and Lytle, 1985). In the case of library bibliographic records, form of expression and concept or abstraction (topic) have been similarly intermingled (Miller, 2000; Crowston and Kwasnik, 2003). However, many sources of guidance are available for encoding information related to form of expression or genre, with several of the most prominent ones listed in the Library of Congress "Source Codes for Genre" (US Library of Congress, n.d.d). MARC 21 (US Library of Congress, 2010) uses fixed-length fields (006-008) for designating forms of material, has a field for Index Term – Genre/Form (655), and recently added several fields in the 300 range related to form of expression: Content Type (336), Media Type (337), Carrier Type (338), Form of Work (380), Other Distinguishing Characteristics of Work or Expression (381), and Medium of Performance (382).

1.8 Concept or abstraction

For several centuries, librarians and other information professionals have been developing and refining systems to represent the concepts and abstractions associated with target information objects. The representation systems have often taken the form of nomenclatures, controlled subject headings, thesauri and, more recently, ontologies (see below). When making use of such a controlled vocabulary, it is important to be aware that the resulting data elements (instances of the controlled vocabulary terms) are likely themselves to serve as TDOs that require contextual information in order for future users to adequately make sense of them. When reading the cause of death on a death certificate, for example, many users would benefit from access to information about the formal nomenclature used to generate the wording used for cause of death as well as the prevailing conventions (e.g. terms that were systematically avoided in order to avoid social stigmas) for applying that nomenclature in such cases (Bowker and Star, 1999). At a minimum, a repository will often be well served by either preserving instances of the nomenclature documentation over time or ensuring that future users will have ready access to the nomenclature documentation from other sources. Once again, this highlights the importance of treating metadata not only as a set of access terms for discovering items, but also as a source of contextual information for making sense of an item once it is discovered.

1.9 Relationship

No formal information system can represent or elaborate all of the relationships that may hold between entities. Instead, small subsets of particularly salient relationships are encoded. Thesauri have traditionally expressed three primary types of relationships: equivalence, hierarchical and associative (ISO 2788, 1986). There are innumerable other types of relationships that can hold between entities (e.g. ancestral, emotional, logistical, causal, temporal, polyhierarchical). Entity-relationship models have long been used to represent relationships of many types, which have generally been implemented using relational databases. Within computer science, the term "ontology" is used to describe data models that accommodate and define an arbitrarily complex set of relationships between entities, concepts, classes or elements. One of the widely proclaimed advantages of the Semantic Web is its support for the definition, tagging and sharing of distributed and often emergent relationships between digital objects and their constituent elements. This could enable unprecedented opportunities for flexible description and interchange of digital information. However, it also raises serious risks for long-term preservation of contextual information, whenever the information characterizing and explaining the relationships that pertain to digital objects is maintained by an institution or individual that does not have the interest or capacity to maintain access to the relationship information over time. RDA (American Library Association, 2008) provides detailed guidance on assigning various types of "relationship designators."

In order to make effective use and sense of a digital object, it can be important to differentiate and provide separate information about the function (purpose), organization (high-level agent) or role responsible for its creation and use, and "personal provenance," i.e. particular individuals involved (Hurley, 1995). Several detailed taxonomies are available for job roles and occupations, including the ERIC Thesaurus, North American Industry Classification System (2007), O*NET Content Model, O*NET-SOC Taxonomy (2009), and Standard Occupational Classification System (2000). METS, Interoperability of Data in E-commerce Systems (INDECS) (Rust and Bide, 2000), the Reference Model for an Open Archival Information System (OAIS) (ISO 14721, 2003) and InterPARES (Long-Term Preservation of Authentic Electronic Records, 2002) all elaborate roles of agents. MARC 21 (US Library of Congress, 2010) includes numerous fields that can be used to identify relationships between the items being cataloged and other resources, as well as allowing for a relator term, which "describes the relationship between a name and a work"; the Library of Congress provides a detailed MARC Value List for Relators and Roles (US Library of Congress, 2003). In his investigation of collection relationships, Heaney (2000) also provides a list of "Types of Agent-Object Relationship." Particular types/genres of objects or purposes may require the designation of further roles. For example, an educational video could include, among other roles, actors and actresses, expert consultant in a video project, director, and producer. The large amount of text that is included in the "credits" of most contemporary Hollywood movies is testament to the numerous roles (and names of associated agents) that one might identify. The Union List of Artist Names (ULAN) elaborates several dozen roles for use in a Person/Corporate Body record. In common language, we often treat roles as attributes of the agent him/herself. As a matter of descriptive convention, roles and job titles also often appear within the metadata associated with a particular agent.