

# Taking Context Seriously: A Framework for Contextual Information in Digital Collections

UNC SILS Technical Report 2007-04

October 18, 2007

**Christopher A. Lee**

*School of Information and Library Science, University of North Carolina, Chapel Hill, NC 27599. Phone: 919-962-7024, Fax: 919-962-8071. Email: callee@ils.unc.edu*

## ABSTRACT

Future users of digital objects will likely 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 contextual information. This paper provides an analysis of context, distinguishing three main ways in which that term has been used within the scholarly literature. I then discuss contextual information within digital collections. I present 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. The paper then discusses existing standards and guidance documents for encoding information related to the nine classes of contextual entities, and it concludes with a discussion of potential implications for descriptive practices through the lifecycle of digital objects.

*“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 are taking 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. Future users will likely 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 once they have them. Literature about the curation of digital collections frequently cites the importance of reflecting context or contextual information associated with digital objects.<sup>1</sup> However, there has been relatively little detailed discussion of (1) what “contextual information” means and (2) how curators of digital collections might best create, capture, encode, manage and provide access to contextual information. This paper addresses both questions. It was written as part of the VidArch Project, which is investigating approaches for creating, capturing and preserving contextual information associated with digital video collections; consequently, many of my examples involve digital video.

## 2. What is Context?

Broadly speaking, context is “the circumstances that form the setting for an event, statement, or idea” (*Compact Oxford English Dictionary*, 2005). Context is an inherently relational concept. 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 – as through the eyes of God,

---

<sup>1</sup> For a recent formulation of this position, see Ross (2007).

LaPlace's demon, or someone inside the Total Perspective Vortex (Adams, 1980). Such a conception of context, however, would not be very useful, nor would it reflect the "thrownness" (Heidegger, 1996; Winograd & Flores, 1986), "embodiment" (Lakoff & Johnson, 1999; Dourish, 2001), or "situatedness" (Lindblom & 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. A more human-centered version of the "everything else" definition might at least scale it back to "anything that is not defined as the phenomenon of interest" (Dervin, 1997). Glaser & 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."

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 & Lenat, 1994) to the TE along some particular dimension or for some 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: (a) TE, (b) context of the TE, and (c) things not relevant enough to be considered part of either (a) or (b). Stated another way, the boundary between content and context is "pragmatic, permeable and revisable" (Callon & Law, 1989) and "is continually negotiated and re-negotiated" (Lea, O'Shea & Fung, 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 & Nardi, 2006), practice (Bourdieu, 1977), domain (Hjørland & Albrechtsen, 1995), situated action (Suchman, 1987), scene (Fillmore, 1977; Blum, 2003; Tyler & Evans, 2003), situation (Dervin, 1983), episode and setting.

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

**Context<sub>1</sub>** - 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 (e.g. the discourse within which a statement is embedded; other documents filed in the same category; formal theory within which a concept or statement is to be understood)

**Context<sub>2</sub>** - objective or socially constructed characteristics and conditions of the situation in which a TE is, appears or occurs (e.g. location; temperature; being under water; occurring as part of a traditional ritual<sup>2</sup>; position within the reporting structure of an organizational hierarchy; relative arrangement and orientation of objects; existence and accessibility of other surrounding objects<sup>3</sup>)

**Context<sub>3</sub>** - 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<sub>1</sub>) is about a TE's place within a larger discourse or information system. 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<sub>1</sub>. The second (context<sub>2</sub>) is about the objective or inter-subjectively recognized set of factors surrounding a TE.<sup>4</sup> The third (context<sub>3</sub>) is about the subjective status of a particular agent -- e.g. a user in context-sensitive computing, one ascribing knowledge to a statement in epistemology, a participant in a speech act. This type of context includes not only

---

<sup>2</sup> 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<sub>2</sub>. However, any culture worthy of the name will also manifest itself in various ways through context<sub>1</sub> and context<sub>3</sub>.

<sup>3</sup> To the extent that they serve as "representational artifacts" (Levy, 2001), surrounding objects can serve as both context<sub>2</sub> and context<sub>1</sub>. 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 & Harper, 2002).

<sup>4</sup> One formulation of the "context" of a statement is precisely that portion of context<sub>2</sub> that is not reflected in context<sub>1</sub>. 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.

“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 habits (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 & Sundaram (2007) break down context<sub>3</sub> further as either context of construction or context of interpretation (context for the message transmitter or receiver, respectively). “Common ground” is shared context<sub>3</sub> that allows two or more individuals to understand each other (Stalnaker, 1978; Clark, 1996). The confluence of contexts<sup>3</sup> (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<sub>3</sub> can take the form of (1) A’s own state, disposition, etc. or (2) the state, disposition, etc. of other agents that are relevant to the matter at hand<sup>5</sup>. Borrowing terminology from sociology and economics, we could call (1) “ego context<sub>3</sub>” and (2) “alter context<sub>3</sub>.” Glaser & Strauss’s (1964) concept of “awareness context” is a subset of context<sub>3</sub>: “the total combination of what specific people, groups, organizations, communities or nations know what about a specific issue.”

Closely related to all three types of context is the importance of history and “path dependence” (Liebowitz & Margolis, 2000); the relevant context often includes aspects of previous states, and not just the current state, 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 5 seconds ago is more likely to serve as context<sub>1</sub> for Betty’s current statement than is a statement that Alice made 5 hours ago. In a game of billiards, the context<sub>2</sub> for my next shot is more clearly defined by my opponent’s previous shot than it is by a shot someone took yesterday on a different table. Likewise, the context<sub>3</sub> in which one interprets a new experience is more clearly defined by her recent thoughts and experiences than by a childhood experience. Because it involves an intentional agent, context<sub>3</sub> 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.

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<sub>2</sub>) or something that she has personally experienced or is currently experiencing (her context<sub>3</sub>). Those who encounter T can then make sense of T by using a combination of (1) the contents of T, (2) other related information entities that are at hand (context<sub>1</sub>), (3) characteristics of the current state of things (context<sub>2</sub>), and (4) their own repertoire of knowledge or experiences at the time (context<sub>3</sub>). 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 this dynamic interchange between artifacts, symbols and a surrounding set of conditions or activities. For example, Léger, Tijus & Baccino (2005) distinguish task, visual and semantic context; Ingwersen & 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,<sup>6</sup> databases, library catalogs) also act at the intersection of context<sub>1</sub>, context<sub>2</sub> and context<sub>3</sub>. Many of the conventions for activities that build and maintain our “epistemic infrastructure” (Hedstrom & King, 2006) - e.g. scholarly research, developing inventories of collections - are attempts to surround expressions about aspects of the world (context<sub>2</sub>) with sufficient and appropriate supporting information (context<sub>1</sub>) so that they will be interpreted by future readers or listeners in (at least relatively) predictable ways.

---

<sup>5</sup> 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.

<sup>6</sup> A theory or other system of inter-related background statements/assumptions is a case of context<sub>1</sub> when it is explicitly encoded into an information system (see e.g. McCarthy, 1993; Guha, 1991; Guha & Lenat, 1994), but it can constitute part of context<sub>3</sub> when it is “in the head” of a particular agent or is otherwise brought to bear on his/her interpretations of, understanding of, actions within or noticing of features within the situation at hand.

Archeologists, for example, attempt to thoroughly document the context - defined as "the relationship of artifacts and other cultural remains to each other and the situation in which they are found" (Society for American Archeology, 1996) - before engaging in activities that might disrupt it, such as moving or manipulating artifacts. A "point of sharp contrast between the archaeologist and the looter" is that the latter "does not bother to record" contextual information before removing an object (Sharer & Ashmore, 1987). An ideal set of descriptive information (context<sub>1</sub>) 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<sub>2</sub>). 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<sub>2</sub>). 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, 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" (Rules for Archival Description, 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, Lee, Olson, & Lampe, 2006) For example, a user in 2057 of a video that was uploaded to YouTube in 2007, might wonder why the creator of that video decided to make it so short. It would be helpful to know that the current terms of YouTube (which changed during the course of writing this paper) indicate, "All videos uploaded to YouTube have a 100MB file size limit. The longer the video is, the more compression will be required to fit it into that size. For that reason, most videos on YouTube are under five minutes long and there is a 10-minute length limit for all videos."

### 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 (Shanon, 1990) – 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<sub>1</sub>) in order to capture and maintain relevant aspects of context<sub>2</sub> and context<sub>3</sub>. "Context, in principle, is infinite. The describer selects certain layers for inclusion, and decides which of those to foreground." (Duff & Harris, 2002) As Heraclitus tells us, "It is impossible to step into the same river twice." No two situations (context<sub>2</sub>) 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 & 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 (1) suffer from gaps in understanding, but also (2) 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, Currall, Moss & Stuart, 2004)

Relationships to other digital objects can dramatically affect the ways in which digital objects have been perceived and experienced. For example, in order for a future user to make sense of a digital video object, it could be useful for that user to know precisely what set of surrogates – textual descriptions, titles, captions, and annotations, as well as surrogates that are themselves still or moving images: video segments, keyframes, skims, slide shows, and fast forwards (Wildemuth, Marchionini, Wilkens, Yang, Geisler, Fowler, Hughes & Mu, 2002; Christel, Smith, Taylor & Winkler, 1998) – were associated with that digital video object at a given point in time. For example, those submitting videos to YouTube must “choose at least one category and enter at least one tag to describe the content in your video” (Making and Optimizing your Videos, 2007). It can be important for a future user to know (1) that this was a requirement of the system and (2) what categories and tags the contributor associated with the video. When someone submits a video to YouTube, the system extracts a frame from the middle of the video to present as the storyboard for that video. Submitters can game this system by placing an image in the middle of the video that is not representative of the video as a whole; or they can be unpleasantly surprised to discover that the automatically-generated key frame does not convey the visual information they had intended. Other online video sharing systems allow submitters to identify a poster frame more directly by embedding this into the video file (e.g. in a QuickTime or MPEG-4 container), or identifying a separate image file. In Open Video – a large collection of digital videos at the University of North Carolina at Chapel Hill – representative surrogate images (storyboards, keyframes and fast forwards) are generated through a combination of automatic processing and manual selection. A future user who was not aware of (1) the specific surrogate images that appeared in an online collection and (2) how those surrogates were generated, may fail to understand the expression, use and perception of the video 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 & 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, Schneider & Dougherty, 2003). Ranking and classification of individual objects can also influence the ways in which they are presented to and experienced by users. In YouTube, for example, some contributors have been accused of creating false subscriptions for the sole purpose of raising their number of channel views and subscribers, or using software to “view” a video many times, thus raising their visibility within the system. Understanding these factors could be essential for future users to understand how a given video 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 recent descriptive and metadata standards, which specify classes, entities, elements, properties and attributes that can serve as contextual information for a target object.<sup>7</sup> Encoded Archival Context (EAC) is a recent effort specifically to formalize contextual information related to archival materials (Ad Hoc Encoded Archival Context Working Group, 2004).

---

<sup>7</sup> Several that have informed this paper are: Data Dictionary–Technical Metadata for Digital Still Images (National Information Standards Organization, 2006); Describing Archives: A Content Standard (DACCS) (2004); Data Documentation Initiative (DDI), Version 3; Encoded Archival Description (EAD) (2002); Metadata Encoding and Transmission Standard (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); Metadata Object Description Standard (MODS); MPEG-21; PReservation Metadata: Implementation Strategies (PREMIS) (PREMIS Working Group, 2005); RUCore Data Model (Weber & Favaro, 2007); Rules for Archival Description (RAD) (2003); Text Encoding Initiative (TEI) (Sperberg-McQueen & Burnard, 2002); PRISM (IDEAlliance Publishing Requirements for Industry Standard Metadata, 2006); VRA Core 4.0 (2007); and Categories for the Description of Works of Art (CDWA) (Baca & Harpring, 2006). The Perseus Project allows for searching on place, person, date or date range, and 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 & Jones, 2006; Smith, 2002; Smith & Crane, 2001)

Literature about information needs in various user populations suggests the value of identifying contextual elements related to the objects in a collection. According to Bearman (1989), "In our culture the fundamental dimensions of reality are space, time, subject, action, object, form and function. The dominance of one or another of these dimensions in any particular construct of the world is a reflection of perspective of the user. User queries are intended to retrieve those representations of realities which have something in common along one or more selected dimensions." The particularly important role of proper names (people and places) in the research of humanities scholars is reported in Wiberley & Jones (1989) and Buchanan, Cunningham, Blandford, Rimmer & Warwick (2005), who also report the use of chronological periods. In her study of abstracts in history, Tibbo (1993) identified the importance of time, place, historical players and events. Bates, Wilde & Siegfried (1993) (Bates, 1996) analyzed natural language statements and queries of humanities scholars, finding that they 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 & Johnson (2001) identified the following elements: proper names, dates, places, subject, form, and, events. In queries to ABC-Clio, Yi, Beheshti, Cole, Leide & Large (2006) found instances of queries based on historical events, people, and regions. Chin & 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 & 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. It is important to recognize that contextual information is not only useful for navigation and discovery of relevant items but also for understanding and making sense of an 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."

Several high-level conceptual models or ontologies provide useful building blocks. These include the ABC Ontology (Lagoze & Hunter, 2001); CIDOC CRM (ISO 21127, 2006); Cyc (Guha & Lenat, 1994); SUMO (Niles & Pease, 2001); Large Scale Concept Ontology for Multimedia (LSCOM) (Naphade, Smith, Tesic, Chang, Hsu, Kennedy, Hauptmann & Curtis, 2006); Functional Requirements for Bibliographic Records (FRBR) (IFLA Study Group, 1988); and Functional Requirements for Authority Data (IFLA Working Group, 2007). Based on the "garbage can model" of decision making in organizations, context can be represented as a particular combination of problems, solutions, participants and choice opportunities (Cohen, March & Olson, 1972). The Reference Model for an Open Archival Information System (OAIS) (ISO 14721, 2003) has served as an influential conceptual framework and source of terminology in the literature on long-term digital preservation (Lee, 2005). The OAIS 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<sub>1</sub>), similar characteristics of the world (context<sub>2</sub>), and commonality with the resources' creators, in background experience, perspective, or knowledge (context<sub>3</sub>). This means that the archives must either capture or create additional documentary material in order to help users make meaning of what I will call a given target digital object (TDO). Within the OAIS information model, there are specific categories of Preservation Description Information that are called Context Information and Provenance Information. However, there also can be valuable sources of information related to context that reside in other parts of the OAIS, including Finding Aids; other Associated Descriptions, such as indexes, annotations, surrogates and new aggregations (Access); and use history data, which falls within Data Management Data.

There are several other relevant standards that define and potentially support the interactions between a diverse set of entities across their life cycle. 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 & Bide, 2000; Wang, Damartini, Wragg, Paramasivam & Barlas, 2005). The Lightweight Directory Access Protocol (LDAP) based on the x.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).

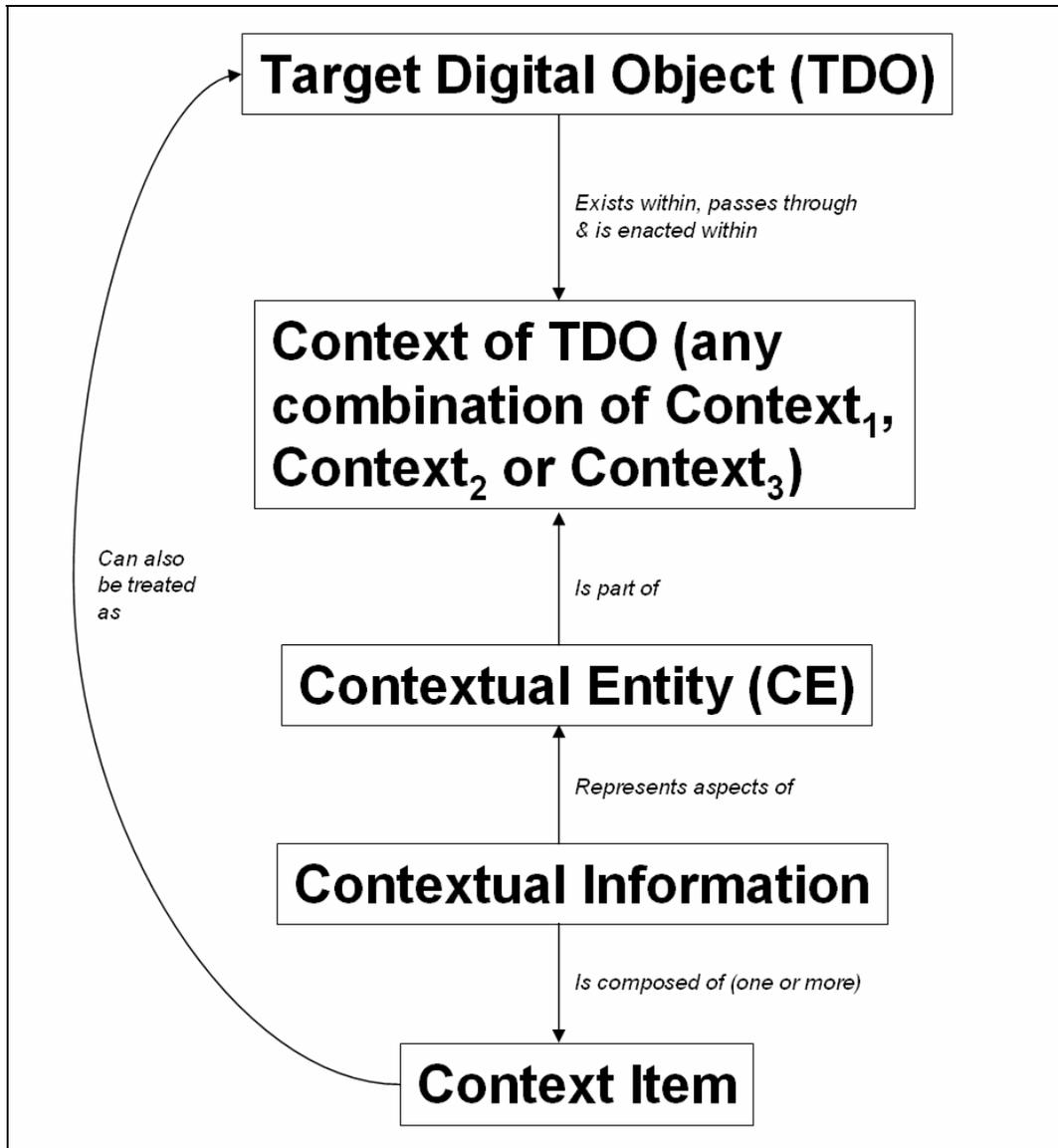
Contextual information is often characterized by recursive relationships. A component of information that was created or captured to provide contextual information about a TDO can often itself become a TDO, about which one may create, capture, and manage additional contextual information (see Figure 1). Like most recursive relationships, there is always a practical limit to the recursion. Holdsworth & 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, 2003) and is often conveyed through individuals and communities of practice (Birnholtz & 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 I present 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 & Theimer, 1994) which attends to the current location and configuration of a mobile device.

### **5.1 Basic Definitions**

In my framework, 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.



**Figure 1 - Relationships between Target Digital Object, Context, Contextual Information, and Context Item**

Throughout its existence, a TDO could pass through innumerable contexts<sup>2</sup>, 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 lifecycle 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 an archive (see Figure 2).

## 5.2 Classes of Contextual Entities

The following table presents the nine classes of contextual entities.

**Table 1 - Nine Classes of Contextual Entities**

<b>Object</b>	a bounded discrete entity that can (1) be characterized as having one or more properties or states, (2) persist across multiple points in time and place, (3) be uniquely identified, (4) interact with other objects, and (5) be acted upon by an agent
<b>Agent</b>	an entity that can carry out actions
<b>Occurrence</b>	a characterization, for a given span of times and places, of either (1) the state of a set of entities or (2) their interaction(s)
<b>Purpose</b>	mandate, norms, values, intention, rules, standards, virtues, or functions to which agents can (1) advance or conform; (2) attempt to advance or conform; (3), hope to advance or conform; or (4) perceive/expect entities (or sets of entities) to advance or conform
<b>Time</b>	“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)
<b>Place</b>	a designated point or region in space
<b>Form of expression</b>	a particular way of expressing ideas or information
<b>Concept or Abstraction</b>	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)
<b>Relationship</b>	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

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).<sup>8</sup> 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.

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, Krafft, Payette & Jesuroga (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 United States (Ferro, Gerber, Mani, Sundheim & Wilson, 2005), and “the White House” could be either an object, agent, place, or purpose.

### 5.2.1 Object

An object is a bounded entity that can (1) be characterized as having one or more properties or states, (2) persist across multiple points in time and place, (3) be uniquely identified, (4) interact with other objects, and (5) 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

<sup>8</sup> WordNet, Version 3.0 (2006) defines the “agentive role” as “the semantic role of the animate entity that instigates or causes the happening denoted by the verb in the clause.”

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 & 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)<sup>9</sup>, 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, Palmer & Robinson, 2000). Collections are usually defined by relationships (see section 5.9 below) between one or both of the following: (1) the objects that make up the collection (e.g. similarity in topic or documentary form) or (2) 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 the collection and what level they consider to be an object within a collection. One advantage of treating a collection as a separate 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 (1) some degree of intentionality (see purpose below) and (2) 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 currently defines agents more narrowly as either persons or groups that “have a role in the lifecycle of a resource.” (Wilson & Clayphan, 2004) 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, Schwartz, Brandherm, Schmitz & von Wilamowitz-Moellendorff, 2005)

The boundary between objects and agents is one of convention.<sup>10</sup> 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 REL 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 & Nardi (2006) identify six different “forms of agency”: natural things, cultural things, natural nonhuman living beings, cultural nonhuman living beings, human beings and 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,

---

<sup>9</sup> See, for example, the distinctions in FRBR between work, expression, manifestation and item; and in PREMIS between intellectual entities and digital objects.

<sup>10</sup> 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).

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 (1) the state of a set of entities or (2) 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, Castaño, Ingria, Saurí, Gaizauskas, Setzer, A., & Katz, 2003). A process is “a naturally occurring or designed sequence of changes of properties or attributes of an object or system” (Process, 2007) – 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 & Hunter, 2001) in which “the state of the world changes” (Matuszek, Cabral, Witbrock, & DeOliveira, 2006). Events tend to be “countable” (Mourelatos, 1978), whereas processes can reach across indefinite spans of space and time, often making them difficult to count. As Dewey says, “every occurrence is a *concurrence*” of elements in an ongoing flow of existence, i.e. “it is inherently characterized by something from which and to which” (1931), so any characterization of something as a completely discrete event is a dramatic 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”) could 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). 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 particularly relevant to digital archives are: those in the world that might be related to the objects preserved (e.g. Hurricane Katrina) and things that happen to the objects themselves (e.g. transformation from one file format to another, transfer of custody, use or annotation).

### 5.2.4 Purpose

The purpose category includes mandates, norms, values, intentions, rules, standards, virtues, or functions which agents can (1) advance or conform to; (2) attempt to advance or conform to; (3), hope to advance or conform to; or (4) perceive/expect entities (or sets of entities) to advance or conform to.<sup>11</sup> 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, Cumming & McKemmish, 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

---

<sup>11</sup> 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.

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 hierarchical or nested relationships with other functions (Australian Governments' Interactive Functions Thesaurus, 2005; Robinson, 1997).

Objects cannot have purposes, but they very 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, Froehlich & Dourish, 2005). Such embedded normative, political and organizational purposes are often not apparent simply from interacting with the objects, and 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 & Star (1999) call "infrastructural inversion" and what DeSanctis & 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" (Leifer, 1991) 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." The most straightforward case is a precise point in time which can be represented as a time and date. However, there is a myriad of other possible temporal units and expressions, which TIMEX2 attempts to accommodate (Ferro et al, 2005), 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"). 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. 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 often 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, Mandreoli & Tiberio, 2005).

### 5.2.6 *Place*

A place is a designated point or region in space. Place information can often 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; 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 & 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 & 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, Abowd & Salber (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. a theater, farm or drive-thru lane). However, in their discussion of “locales,” Fitzpatrick, Kaplan & Mansfield (1996) point out that place/purpose mappings are often quite complicated and dynamic.

### 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. This can include genre, terminology (e.g. the lingo of a particular group), 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 & 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) The ABC Ontology distinguishes between an “actuality” and an abstraction (Lagoze & 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? Should we consider a fictional character, such as Napoleon Dynamite, to be just an idea (abstraction), or as 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 is going to provide definitive resolutions to such questions. Instead, a framework for contextual information can foreground the importance, in some cases, of capturing, fixing and preserving digital representations of some 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 the digital objects lifecycle?

### 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. 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, Chapman & Cheney, 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, Khanna & Tan, 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. The case for placing relationships within the foreground of investigation gained wide visibility with Granovetter’s work on the strength of weak ties (1973). By the mid-1990s, there was a well-established set of methods for approaching practically any social or socio-technical phenomenon as a network of connections, interactions, or resource flows – many of the fundamental methodological elements being synthesized and explained by Wasserman & 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 (1) one or more agents and (2) 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 George W. Bush. This role and its related purposes preexist his term of office and they will persist after he leaves office.<sup>12</sup>

## 6. Implications for Descriptive and Curatorial Practice

Consider the following scenario: The year is 2057. You use a search tool to locate and download a 3-minute video file. There is no information associated with the file to tell you anything about it. Through a form of file recognition or transformation that you do not (nor do you care to) know anything about, your viewing device is able to render the old file, and you watch the video. A skinny young man with unruly hair, wearing black boots, blue jeans, glasses and a white t-shirt with the words “Vote for Pedro” walks onto a stage. Someone off-stage turns on music that sounds like it might be from around the late 20th century, using some sort of broadcast equipment that looks completely unfamiliar to you. The skinny young man then proceeds to dance in front of an audience of what appear to be teenagers. You are intrigued but have no idea what you have just watched.

---

<sup>12</sup> Agents and purposes often co-evolve. Every person who has served as President of the United States has been changed by the experience; and the functions/purposes of the presidency have often changed significantly as a result, in part, of a different person holding the office (some changes in the *de jure* functions/purposes, but even more changes in the *de facto* functions/purposes).

You download and view another video, which is about 2 minutes long. It begins with a view of a beach from what appears to be a tropical resort somewhere. A male voice (perhaps the voice of the man holding the video camera) states in a calm but concerned tone, “It’s coming in. It’s coming again.” The camera shakes and shifts a lot. There is a rapid exchange of brief statements (some in English and some in a language that you don’t recognize) that appear to be related to whatever it is that is “coming.” The movement of the camera gets even more erratic. About 30 seconds into the video, it becomes apparent that water is rushing into the building, and there are sounds of screaming. Throughout the rest of the video, the rush of water into the building gets increasingly more dramatic. Two people and countless pieces of furniture are swept away. The video ends with both the image and the audio breaking up.<sup>13</sup>

What more would one want or need to know about the videos described above, in order to make sense of them? Many viewers of the videos in 2007 will have a very good idea of what they are seeing, for two fundamental reasons. First, they have recent memories (either first-hand or through exposure to popular discussion and media coverage) of the phenomena upon which the videos are based. Second, they are likely to have access to associated documentary materials that can fill in the gaps. The first condition will not be satisfied in 2057, and the second condition will only be met in 2057 if archives, libraries and other institutions take responsibility for preserving associated contextual information.

## 6.1 Context and Contextual Information in Digital Collections

Custodians of digital collections must decide: (1) what aspects of the digital objects’ creation and use environment are important enough to warrant capture, documentation, and preservation over time; (2) given limited resources and available technology, which of those aspects of context can reasonably be preserved; and (3) how to carry out the preservation (and often the creation and capture) of the contextual information.

Rather than treating each item as a discrete entity, archival theory and practice suggest that digital objects should be managed, preserved and presented to users in a way that reflects 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” (Curral, Moss & Stuart, 2004). Archivists attempt to reflect this form of agency by retaining the “original order” of materials. Arrangement and description of archival materials involves: “gathering information about the person, organization, or group that created, accumulated, assembled, or used a group of records”; “identify[ing] the function or roles records were created to support”; “identify[ing] recordkeeping practices evidenced in the records”; and “identify[ing] significant events or developments to which the records relate.” (Roe, 2005) Several authors within the literature on archival description have argued strongly for the importance of contextual information – about entities and conditions associated with document creation – as distinct from bibliographic information, which describes characteristics of the document itself (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, including what they decide not to keep or describe. Margaret Hedstrom argues for “concepts, tools, and processes that enable 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” (2002).

## 6.2 Representing and Preserving Contextual Information

Throughout its existence, a TDO could pass through innumerable contexts<sup>2</sup>, 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 lifecycle 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 an archive (see Figure 2).

---

<sup>13</sup> The first video, “Napoleon Dynamite Dance Scene” (<http://www.youtube.com/watch?v=ixsZy2425eY>), was posted to YouTube on November 25, 2005 by BloodTempest. YouTube ranks it as the 34th most viewed video (8,310,486 views). The second video, “Tsunami - Phuket” ([http://www.archive.org/details/tsunami\\_phuket](http://www.archive.org/details/tsunami_phuket)), was created in 2004. It is available from the Internet Archive, which lists it as the most downloaded item in its Moving Image Archive (639,118 downloads). Numbers and ranks are from June 7, 2007.

One of the complications of digital objects is that they do not reflect the exact same properties in all contexts – i.e. they do not conform to Leibniz’s Law (Allison, Moss & Stuart, 2004). Two situations are worth noting. First, at a given point in 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. In both cases, one may either consider (1) each distinct instance of the digital object to be a new item worthy of its own identity and management, or (2) 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 lifecycle (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. Within the language of the Functional Requirements for Bibliographic Records (FRBR), 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 (not directly addressed in FRBR) unique projection onto computer hardware of an existing item (e.g. identical bitstream mirrored in a second location or cached for quick access).

Descriptive standards for libraries and archives have historically been much more thorough in specifying how to describe the things within a collection (e.g. bibliographic elements about books) 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. Encoded Archival Context (EAC) is a recent effort to formalize contextual information related to archival materials (Ad Hoc Encoded Archival Context Working Group, 2004). The Research Support Libraries Programme (RSLP) has also produced several guidance documents for collection-level description (e.g. Powell, Heaney & Dempsey, 2000). Many other recent descriptive and metadata standards – as well as several high-level conceptual models or ontologies – specify classes, entities, elements, properties and attributes that can serve as contextual information for a target object.<sup>14</sup>

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) (2002); General International Standard Archival Description (ISAD(G)) (Ad Hoc Commission on Descriptive Standards, 2000); Rules for Archival Description (RAD) (2003); and DACS – include elements related to the context of creation of the finding aids themselves. Light & Hyry (2002) argue for the incorporation of further contextual information throughout the lifecycle of the finding aid.

Building up a network of descriptive information about the nine types of entities is different from simply trying to take a snapshot of the context<sub>2</sub> 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. where she was born, names of her parents, where she went to school, 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 also support browsing and serendipitous discovery (e.g. “Oh, so she also wrote that other book (X). Let’s find out more about X. Oh, X was based on her experiences working at Y. Let’s find out more about Y. Oh, George also worked at Y. I wonder if he wrote anything...”).

### 6.3 Contextual Information about the Nine Classes of Contextual Entities

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

---

<sup>14</sup> See footnote 7 for a more detailed elaboration of those that have informed my development of the contextual information framework.

### 6.3.1 *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 (An Introduction, 2006), to the Categories for the Description of Works of Art (CDWA) for art and material culture (Baca & Harpring, 2006), to the relatively institution-specific conventions for representing archeological artifacts (Snow, Gahegan, Giles, Hirth, Milner, Mitra & Wang, 2006). No repository will fully adopt all of these standards. However, in cases when collections of 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) (Report Concerning Space Data System Standards, 1992), and its intended successor XML Formatted Data Unit (XFDU), for space and terrestrial data; MPEG-21 for video (ISO/IEC 21000-2, 2005); and IMS Learner Information Package Information Model Specification (Smythe, Tansey & Robson, 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, De Kooning Van de Walle, 2005). The Metadata Encoding and Transmission Standard (METS) (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 objects 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 & Lee, 2002).

### 6.3.2 *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. The Library of Congress has maintained the Name Authority File for this purpose, and has joined with other major libraries from the English-speaking world to develop the Anglo-American Authority File (AAAF). Other rich sources of biographical and name information are the Union List of Artist Names (ULAN) and Oxford Dictionary of National Biography. 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 & Clayphan, 2004). A project within the International Organization for Standardization (ISO), which began in Fall 2006, is aiming to develop the International Standard Party Identifier (ISPI) (ISO Project 27729). International Standard Archival Authority Record for Corporate Bodies, Persons, and Families (ISAAR(CPF) (ICA Committee on Descriptive Standards, 2004) and Encoded Archival Context (EAC) are two rich sources of guidance on the types of information one might hope to provide about agents.

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 & 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 (Hazelton, 2007) 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 & Miller, 2007). OpenID defines “eight commonly requested pieces of information” about individuals (Hoyt, Daugherty & Recordon, 2006). Another source of guidance is the work on user modeling, including the General User Model Ontology (GUMO) (Heckmann, Schwartz, Brandherm, Schmitz & von Wilamowitz-Moellendorff, 2005), the User Modeling Markup Language (UserML) (Heckmann & 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 U.S. Bureau of Labor Statistics. The ERIC Thesaurus also includes a category devoted to “occupations.”

### 6.3.3 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 & Gruninger, 2005); extension and application of the Unified Modeling Language (Penker & Eriksson, 2000); XML Process Definition Language (2005); and the Business Process Modeling Notation Specification (White, 2006). 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, Castaño, Ingria, Saurí, Gaizauskas, Setzer, A., & Katz, 2003); and the HEML schema includes elements for location, time, persons, and evidence for the event (Robertson, 2006). PRISM defines metadata fields “to provide information about an event pictured in the image or contributing to the image” (IDEAlliance Publishing Requirements for Industry Standard Metadata, 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, Demartini, Wragg, Paramasivam & Barlas, 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, Beecroft, Johnson & Ward, 2006).

### 6.3.4 Purpose

Functions often have hierarchical or nested relationships with other functions. Two sources of detailed guidance for representing functional entities and their relationships are the Australian Governments' Interactive Functions Thesaurus (2005) and Keyword AAA (Robinson, 1997). An effort is also currently underway to develop an International Standard for Describing Functions (ICA-ISDF) associated with archival materials (Sibille, 2007). 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.

### 6.3.5 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, Gerber, Mani, Sundheim & Wilson, 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, Larson & Buckland, 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, Mandreoli & Tiberio, 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.

### 6.3.6 *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” (2004). A well-established set of conventions for encoding locations as coordinates is available in the Department of Defense World Geodetic System 1984 (2000), which is supported by vCard and the geo microformat (Çelik, 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, and the Getty Thesaurus of Geographic Names (TGN). EAC distinguishes geographic name (geog) or a jurisdictional area (juris). The creation of geo-referenced data is increasing dramatically, not only through direct entry into dedicated geographic information systems (GIS), but also captured by devices such as digital cameras or later assigned by users as “geotags.” Conventions for identifying place will be important for supporting the interoperability and reuse of the place data. 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 & Hanson, 1984). Although it is closely connected to specific geographic localities, “nationality” (Harpring, Beecroft, Johnson, & Ward, 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.

### 6.3.7 *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 & Lytle, 1985). In the case of library bibliographic records, form of expression and concept or abstraction (topic) have been similarly intermingled (Miller, 2000; Crowston & 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.”

### 6.3.8 *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 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 user cases 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.

### 6.3.9 *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.

In order to make effective use and sense of a digital object, it can be important to differentiate and provide separate information about (1) the function (purpose), organization (high-level agent) or role responsible for its creation and use, and (2) “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 (2002), O\*NET Content Model, O\*NET-SOC Taxonomy (2006), and Standard Occupational Classification System (2000). METS, Interoperability of Data in E-commerce Systems (INDECS) (Rust & Bide, 2000), the Reference Model for an Open Archival Information System (OAIS) (ISO 14721, 2003) and InterPARES all elaborate roles of agents. MARC 21 Concise Format for Bibliographic Data (2006) allows for a relator term, which “describes the relationship between a name and a work”; and the Library of Congress provides a detailed MARC Value List for Relators and Roles (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.

## 6.4 Capturing Contextual Information Throughout the Life of a Digital Object

How much does one need to know about the social, psychological, technical, and organizational context in which a digital object was originally created and used? How, when and why should particular elements of contextual information make their way into a repository? There are practical limitations to how much contextual information can be captured, but 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 & 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 proper 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 & Halverson, 1998)

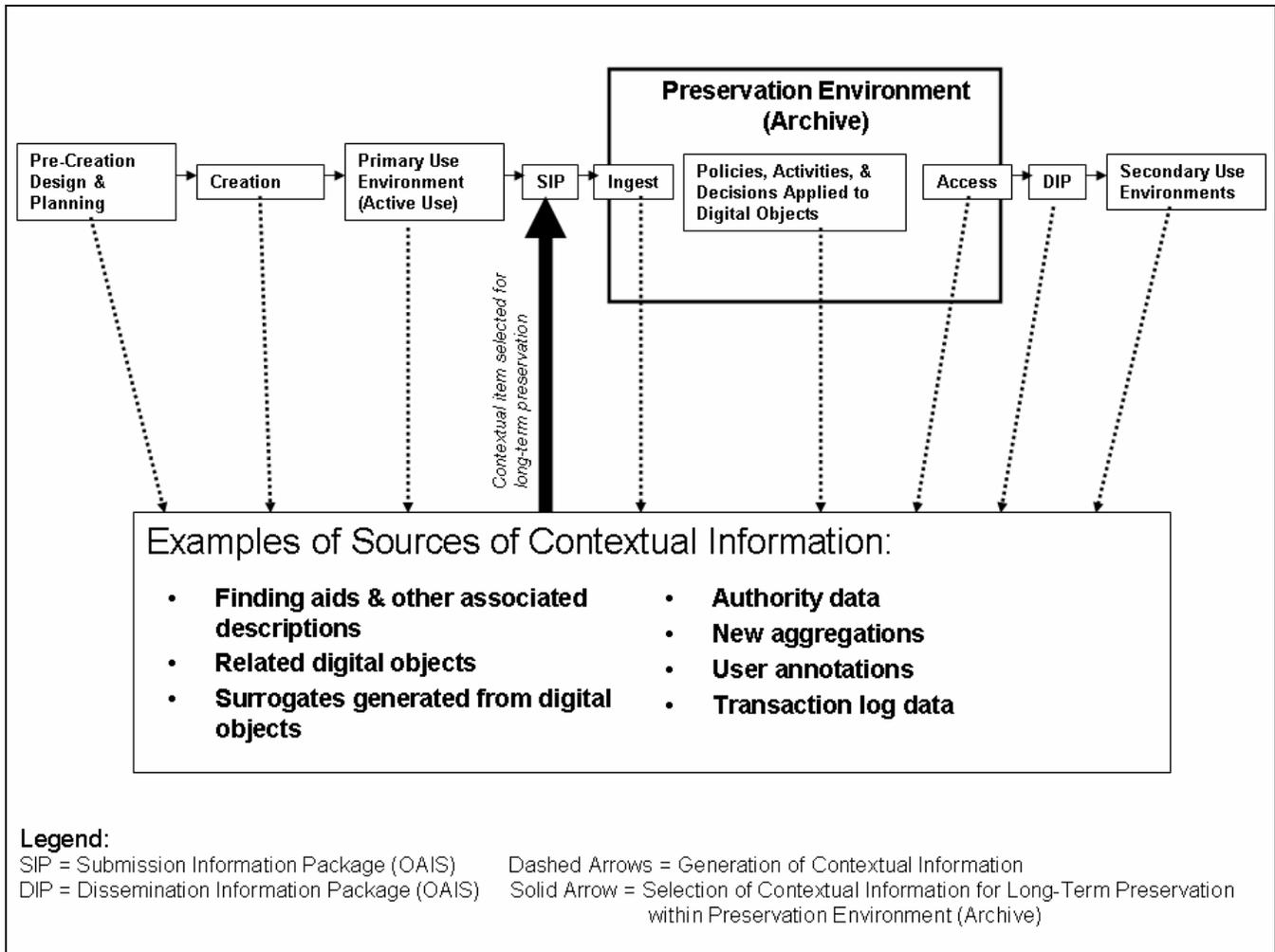
Existing approaches for creating and capturing contextual information associated with TDOs within collections have usually attended to particular, relatively discrete points in the information lifecycle, for example: creation; declaration or endorsement (e.g. signing a document as an official record or designating it to be an official publication of a government agency); 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.

Rather than relying solely on archivists, librarians and other staff within one’s own repository to create contextual information from scratch during one relatively isolated accessioning process, TDOs will ideally “pick up contextual information along with way.” Reference collections within repositories can serve as sources of contextual information (Buckland, 2007), as well as supplemental sources created outside repositories, such as “government directories, government information services, and telephone directories...business 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. Figure 2 presents creation, capture and preservation of contextual information, within a set of broad stages developed in the DigCCurr project (Lee, Tibbo & Schaefer, 2007). 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, Baker & Halkola, 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). The generation of important contextual information often does not stop at the point of creation. 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 & Ashmore, 1987).<sup>15</sup> “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.” (Ad Hoc Commission on Descriptive Standards, 2000) 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 over time, e.g. the ingestion, deletion, migration, or dissemination of a target digital object or associated digital objects.

---

<sup>15</sup> The concept of provenance – both as it is discussed in literature of the archival profession (Bearman & Lytle, 1995), and as conveyed in more recent guidance documents (e.g. OAIS, PREMIS, METS) – spans agents (who create it, who changed it) and occurrences (what occurrences created it, what occurrences changed it). Provenance is a characteristic not “of entities but of relationships between entities” (Hurley, 1995).



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

Another potential source of contextual information is use data, particularly if it is “collected from users in an ongoing program” (Conway, 1986). Many libraries and archives collect information about the use of their materials. Both ethical and administrative issues, however, 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. Counting Online Usage of Networked Electronic Resources (COUNTER) and the Developing Archival Metrics Project (MacKay & 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 & Plum, 2006).

There is great promise in using automated or semi-automated techniques to generate contextual information, which can then be presented to users and potentially ingested into the Archive in order to be preserved along with the TDOs. This can include named entity recognition or metadata extraction based on machine learning, natural language processing or recognition of elements embedded into particular file formats (Ciravegna, Chapman, Dingli & Wilks, 2004; Crane & Jones, 2006; Hu, Cao, Meyerzon, and Zheng, 2005; Kim & Ross, 2006; Thoma, Mao & Misra, 2005; Saurí, Knippen, Verhagen & Pustejovsky, 2005); developing collections of facts about known entities using data available on the Web (Shah, Schneider, Matuszek, Kahlert, Aldag, Baxter, Cabral, Witbrock & Curtis, 2006); or inferring the author of portions of text within a collection of threaded messages (Hoffmann, 2007). Software can also support higher-level inferences about contextual information, such as using patterns in the co-occurrence of dates and places within documents to identify events within a

collection (Smith, 2002) or using machine learning to make inferences about temporal relations between events (Mani, Verhagen, Wellner, Lee & Pustejovsky, 2006). Natural language processing techniques could generate suggestions about who authored a particular digital object, when this is not already known, and may even support inferences about the intended sentiment or emotional state of the author (Yu, 2006), or degree of social influence of a particular agent represented in a collection of documents (Fader, Radev, Crespin, Monroe, Quinn & Colaresi, 2007). Such inferred contextual information (potentially confirmed, further described or selected by a human curator) could then be ingested into the Archive 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 (1) discrete, localized and ephemeral, thus not well-documented elsewhere, (2) subtle or complex enough that they would be difficult to identify or disambiguate using automation later, (3) fundamental bridging mechanisms to collections that reside elsewhere, or (3) so fundamental to the collecting mission of the archive that they warrant focused, narrative description.

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). "While a collecting institution may always have an apparent advantage in having access to the real thing', 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 & Trant, 1998) Some users will be able to bring unique information to the archive about aspects of the context<sub>1</sub>, context<sub>2</sub>, and context<sub>3</sub> 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 cycle 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 & 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 (Phelps & Wilensky, 2000; Thiel, Brocks, Frommholz, Dirsch-Weigand, Keiper, Stein & Neuhold, 2004), 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, Gregory, Kuo, Wackerow & Nelson, 2007).

Preservation of contextual information does not always require direct custody. An archive that has TDOs for which it hopes to provide further contextual information can enter into (in the terms of the OAI) cooperating, federated, or shared resource arrangements with other trustworthy archives, who are responsible for managing the contextual information. For purposes of access, an archive can provide links to context items within other archives, 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, Zavalina & Mustafoff, 2007) 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 not only provide access to digital objects but also the means to make meaningful use and sense of the digital objects long into the future.

## 7. 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.

## 8. ACKNOWLEDGMENTS

This work has been supported through NSF Grant # IIS 0455970. Thanks to members of the VidArch team and the Curation and Archives Research group at UNC for discussion of issues raised in this document. Jane Greenberg, Jennifer Engleson Lee, Laura Sheble, Paul Solomon and Helen Tibbo provided valuable suggestions on earlier drafts.

## 9. REFERENCES

- Ackerman, M. S., & Halverson, C. (1998). Considering an organization's memory. In CSCW '98: proceedings: ACM 1998 Conference on Computer Supported Cooperative Work, Seattle, Washington, November 14-18 (pp. 39-48). New York, NY: Association for Computing Machinery.
- Acland, G., Cumming, K., & McKemmish, S. (1999). The End of the Beginning: The SPIRT Recordkeeping Metadata Project. Paper presented at the Australian Society of Archivists Conference. Retrieved June 7, 2007, from <http://www.archivists.org.au/events/conf99/spirt.html>
- Adams, D. (1980). *The restaurant at the end of the universe*. New York, NY: Pocket Books.
- Ad Hoc Commission on Descriptive Standards. (2000). *ISAD(G): General International Standard Archival Description* (2nd ed.). Ottawa, Canada: International Council on Archives.
- Ad Hoc Encoded Archival Context Working Group (Ed.). (2004). *Encoded Archival Context Tag Library*. Retrieved June 7, 2007, from [http://www.iath.virginia.edu/saxon/servlet/SaxonServlet?source=/eac/documents/tl\\_beta.xml&style=/eac/shared/styles/tl.xsl](http://www.iath.virginia.edu/saxon/servlet/SaxonServlet?source=/eac/documents/tl_beta.xml&style=/eac/shared/styles/tl.xsl)
- Agre, P. E. (2001). Changing Places: Contexts of Awareness in Computing. *Human-Computer Interaction*, 16, 177-192.
- Alexandria Digital Library Feature Type Thesaurus. (2002). University of California, Santa Barbara. Retrieved June 7, 2007, from <http://www.alexandria.ucsb.edu/gazetteer/FeatureTypes/ver070302/>
- Allen, J. F. (1984). Towards a general theory of action and time. *Artificial Intelligence*, 23, 123-154.
- Allison, A., Currall, J., Moss, M., & Stuart, S. (2004). Digital Identity Matters. *Journal of the American Society for Information Science and Technology*, 56(4), 364-372.
- Austin, J. L. (1962). *How to do things with words*. Cambridge, MA: Harvard University Press.
- Australian Governments' Interactive Functions Thesaurus – AGIFT. (2nd ed.) (2005). Canberra, Australia: National Archives of Australia. Retrieved on June 7, 2007, from <http://www.naa.gov.au/recordkeeping/thesaurus/index.htm>
- Baca, M., & Harpring, P. (Eds.). (2006). *Categories for the Description of Works of Art*. J.Paul Getty Trust. Retrieved June 28, 2007, from [http://www.getty.edu/research/conducting\\_research/standards/cdwa/](http://www.getty.edu/research/conducting_research/standards/cdwa/)
- Bardram, J. E. (2005). The Java Context Awareness Framework (JCAF) – A Service Infrastructure and Programming Framework for Context-Aware Applications. In H. Gellersen, R. Want & A. Schmidt (Eds.), *Pervasive Computing: Third International Conference, PERVASIVE 2005, Munich, Germany, May 8-13, 2005. Proceedings* (pp. 98-115). Berlin: Springer.
- Bates, M. J., Wilde, D. N., & Siegfried, S. (1993). An analysis of search terminology used by humanities scholars: the Getty Online Searching Project Report Number 1. *Library Quarterly*, 63(1), 1-39.

- Bates, M. J. (1996). The Getty end-user online searching project in the humanities: Report no. 6: Overview and conclusions. *College & Research Libraries*, 57, 514-523.
- Bazerman, C. (1988). *Shaping written knowledge: the genre and activity of the experimental article in science*. Madison, WI: University of Wisconsin Press.
- Bearman, D. (1989). *Archival Methods*. Pittsburgh, PA: Archives and Museum Informatics.
- Bearman, D. (1992). Documenting Documentation. *Archivaria*, 34, 33-49.
- Bearman, D. A., & Lytle, R. H. (1985). The Power of the Principle of Provenance. *Archivaria*, 21, 14-27.
- Bearman, D., & Trant, J. (1998). Authenticity of Digital Resources: Towards a Statement of Requirements in the Research Process. *D-Lib Magazine*, 4(6).
- Bearman, D., & Trant, J. (1999). Unifying our cultural memory: Could electronic environments bridge the historical accidents that fragment cultural collections? In S. Criddle, L. Dempsey & R. Heseltine (Eds.), *Information landscapes for a learning society: networking and the future of libraries 3* (pp. 207-234). London, Library Association Publishing.
- Bekaert, J. L., De Kooning, E., & 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*, 10(4), 286-301.
- Birnholtz, J. P., & Bietz, M. J. (2003). Data at work: supporting sharing in science and engineering. In M. Pendergast (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* (pp. 339-348). New York, NY: Association for Computing Machinery.
- Blum, A. (2003). *The imaginative structure of the city*. Montreal: McGill-Queen's University Press.
- Bock, C., & Gruninger, M. (2005). PSL: A Semantic Domain for Flow Models. *Software and Systems Modeling*, 4(2), 209-231.
- Bourdieu, P. (1977). *Outline of a theory of practice* (R. Nice, Trans.). Cambridge, UK: Cambridge University Press.
- Bowker, G. C., & Star, S. L. (1999). *Sorting things out: classification and its consequences*. Cambridge, MA: MIT Press.
- Brack, E. V., Palmer, D., & Robinson, B. (2000). Collection Level Description - the RIDING and Agora Experience. *D-Lib Magazine*, 6(9).
- Brickley, D., & Miller, L. (2007). FOAF Vocabulary Specification 0.9. Retrieved July 15, 2007, from <http://xmlns.com/foaf/spec/>
- Buchanan, G., Cunningham, S. J., Blandford, A., Rimmer, J., & Warwick, C. (2005). Information Seeking by Humanities Scholars. In S. Christodoulakis, A. Rauber & A. M. Tjoa (Eds.), *Research and advanced technology for digital libraries 9th European conference, ECDL 2005, Vienna, Austria, September 18-23, 2005: proceedings* (pp. 218-229). Berlin: Springer.
- Buckland, M. K. (2007). The Digital Difference in Reference Collections. *Journal of Library Administration*, 46(2), 87-100.
- Buneman, P., Chapman, A., & Cheney, J. (2006). Provenance management in curated databases. In *ACM SIGMOD/PODS 2006 electronic proceedings: Chicago, Illinois, USA, June 26-29, 2006* (pp. 539-550). New York, NY: Association for Computing Machinery.

- Buneman, P., Khanna, S., & Tan, W.-C. (2001). Why and Where: A Characterization of Data Provenance. In J. v. d. Bussche & V. Vianu (Eds.), *Database Theory - ICDT 2001: 8th International Conference*, London, UK, January 2001. Proceedings (pp. 316-330). Berlin: Springer.
- Callon, M., & Law, J. (1989). On the Construction of Sociotechnical Networks: Content and Context Revisited. *Knowledge and Society*, 8, 57-83.
- Çelik, Tantek. Geo – Microformats. Retrieved July 11, 2007, from <http://microformats.org/wiki/geo>
- Chin, G., Jr., & Lansing, C. S. (2004). Capturing and Supporting Contexts for Scientific Data Sharing via the Biological Sciences Collaboratory. In *CSCW 2004: computer supported cooperative work: conference proceedings*, November 6-10, 2004, Chicago (pp. 409-418). New York, NY: ACM Press.
- Christel, M. G., Smith, M. A., Taylor, C. R., & Winkler, D. B. (1998). Evolving Video Skims into Useful Multimedia Abstractions. In S. Pemberton & C.-M. Karat (Eds.), *Human factors in computing systems: CHI'98 Conference proceedings: making the impossible possible*, April 18-23, Los Angeles (pp. 171-178). New York, NY: Association for Computing Machinery.
- Ciravegna, F., Chapman, S., Dingli, A., & Wilks, Y. (2004). Learning to Harvest Information for the Semantic Web. Paper presented at the 1st European Semantic Web Symposium, May 10-12, Heraklion, Greece.
- Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.
- Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A garbage can model of organizational choice. *Administrative Science Quarterly*, 17(1), 1-25.
- Cole, C. (2000). Name Collection by Ph.D. History Students: Inducing Expertise. *Journal of the American Society for Information Science*, 51(5), 444-455.
- Compact Oxford English Dictionary of Current English. (3rd ed.). (2005), New York, NY: Oxford University Press.
- Conway, M. E. (1968). How Do Committees Invent? *Datamation*, 14(4), 28-31.
- Conway, P. (1986). Facts and Frameworks: An Approach to Studying the Users of Archives. *American Archivist*, 49, 393-407.
- COUNTER – Online Usage of Electronic Resources. Retrieved June 8, 2007, from <http://projectcounter.org/>
- Crane, G., & Jones, A. (2006). The challenge of virginia banks: an evaluation of named entity analysis in a 19th-century newspaper collection. In *6th ACM/IEEE-CS Joint Conference on Digital Libraries 2006 Opening information horizons: June 11-15, 2006, Chapel Hill, NC, USA* (pp. 31-40). New York, NY: ACM Press.
- Crowston, K., & Kwasnik, B. H. (2003). Can Document-genre Metadata Improve Information Access to Large Digital Collections? *Library Trends*, 52(2), 345-361.
- Currall, J., Moss, M., & Stuart, S. (2004). What Is a Collection? *Archivaria*, 58, 131-46.
- Dawson, F., & Howes, T. (1998). vCard MIME Directory Profile. (RFC 2426).
- DDI 3.0. (2007). DDI Alliance. Retrieved July 5, 2007, from <http://www.ddialliance.org/ddi3/index.html>
- Department of Defense World Geodetic System 1984: Its Definition and Relationships with Local Geodetic Systems, Third Edition, Amendment 1. (NIMA Technical Report TR8350.2) (2000). St. Louis, MO: National Imagery and Mapping Agency.

- Dervin, B. (1997). Given a context by any other name: Methodological tools for taming the unruly beast. In P. Vakkari, R. Savolainen & B. Dervin (Eds.), *Information Seeking in Context* (pp. 13-38). London: Taylor Graham.
- Dervin, B. (1983, May 26-30). 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.
- DeSanctis, G., & Poole, M. S. (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization Science*, 5(2), 121-147.
- Describing Archives: A Content Standard. (2004). Chicago, IL: Society of American Archivists.
- Dewey, J. (1931). *Context and Thought*. University of California Publications in Philosophy, 12(3), 203-224.
- Dewey, J. (1922). *Human Nature and Conduct: An Introduction to Social Psychology*. New York, NY: Henry Holt & Company.
- Dey, Anind K., Gregory D. Abowd, and Daniel Salber. "A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications." *Human-Computer Interaction* 16, no. 2-4 (2001): 97-166.
- Doerr, M., Hunter, J., & Lagoze, C. (2003). Towards a Core Ontology for Information Integration. *Journal of Digital Information*, 4(1).
- Dourish, P. (2001). *Where the Action Is: The Foundations of Embodied Interaction*. Cambridge, MA: MIT Press.
- Duff, W., & Harris, V. (2002). Stories and Names: Archival Description as Narrating Records and Constructing Meanings. *Archival Science*, 2(3-4), 263-285.
- Duff, W.M., & Johnson, C.A. (2001). A Virtual Expression of Need: An Analysis of E-Mail Reference Questions. *American Archivist*, 64(1), 43-60.
- Edmonds, B. (1999). The Pragmatic Roots of Context. In P. Bouquet, L. Serafini, P. Brezillon, M. Benerecetti & F. Castellani (Eds.), *Modeling and Using Context: Second International and Interdisciplinary Conference, CONTEXT'99*, Trento, Italy, September 1999. Proceedings (pp. 119-134). Berlin: Springer-Verlag.
- Encoded Archival Description Tag Library. (2002). Society of American Archivists and Library of Congress. Retrieved July 15, 2007, from <http://www.loc.gov/eat/tglib/>
- ERIC Thesaurus. U.S. Department of Education, Institute of Education Sciences. <http://www.eric.ed.gov/thesaurus>
- Evans, M. J. (1986). Authority Control: An Alternative to the Record Group Concept. *American Archivist*, 49, 249-261.
- Fader, A., Radev, D. R., Crespian, M. H., Monroe, B. L., Quinn, K. M., & Colaresi, M. (2007). MavenRank: Identifying influential members of the US senate using lexical centrality. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (pp. 658-666). Stroudsburg, PA: Association for Computational Linguistics.
- Ferro, L., Gerber, L., Mani, I., Sundheim, B., & Wilson, G. (2005). TIDES 2005 Standard for the Annotation of Temporal Expressions: MITRE Corporation. Retrieved on June 14, 2007, from [http://timex2.mitre.org/annotation\\_guidelines/2005\\_timex2\\_standard\\_v1.1.pdf](http://timex2.mitre.org/annotation_guidelines/2005_timex2_standard_v1.1.pdf)
- Fillmore, C. J. (1977). Scenes-and-frames semantics. In A. Zampolli (Ed.), *Linguistics Structures Processing* (pp. 55-81). Amsterdam: North-Holland.
- Fitzpatrick, G., Kaplan, S., & Mansfield, T. (1996). Physical spaces, virtual places and social worlds: A study of work in the virtual. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*. New York, NY: ACM Press.

- Foot, K., Schneider, S., & Dougherty, M. (2003, October 16-19). The Ethics of/in Web Archiving. Paper presented at Broadening the Band: Association of Internet Researchers Conference, Toronto, Ontario, Canada. Retrieved July 15, 2007, from [http://www.webarchivist.org/031009.ethics-web-archiving\\_ReadOnly.ppt](http://www.webarchivist.org/031009.ethics-web-archiving_ReadOnly.ppt)
- Franklin, B., & Plum, T. (2006). Successful Web Survey Methodologies for Measuring the Impact of Networked Electronic Services (MINES for Libraries). *IFLA Journal*, 32(1), 28-40.
- Gadamer, H. G. (1989). *Truth and method* (2nd rev. ed.). (J. Weinsheimer & D.G. Marshall, Trans.). New York: Crossroad.
- Geographic Names Information System. U.S. Board on Geographic Names. Retrieved June 8, 2007, from <http://geonames.usgs.gov/>
- Getty Thesaurus of Geographic Names. J. Paul Getty Trust. Retrieved June 8, 2007, from [http://www.getty.edu/research/conducting\\_research/vocabularies/tgn/](http://www.getty.edu/research/conducting_research/vocabularies/tgn/)
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.
- Glaser, B. G., & Strauss, A. L. (1964). Awareness Contexts and Social Interaction. *American Sociological Review*, 29(5), 669-679.
- Grandi, F., Mandreoli, F., & Tiberio, P. (2005). Temporal modelling and management of normative documents in XML format. *Data & Knowledge Engineering*, 54(3), 327-354.
- Granovetter, M. S. (1973). The Strength of Weak Ties. *American Economic Review*, 78(6), 1360-1480.
- Greenberg, S. (2001). Context as a Dynamic Construct. *Human-Computer Interaction*, 16(2/4), 257-268.
- Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Grudin, J. (2001). Desituating Action: Digital Representation of Context. *Human-Computer Interaction*, 16(2/4), 269-286.
- Guha, R. V. (1991). *Contexts: A Formalization and Some Applications*. Unpublished PhD Dissertation, Stanford University, Stanford, CA.
- Guha, R. V., & Lenat, D. B. (1994). Enabling agents to work together. *Communications of the ACM*, 37(7), 126-142.
- Guha, R., & McCarthy, J. (2003). Varieties of Contexts. In P. Blackburn, C. Ghidini, R. M. Turner & F. Giunchiglia (Eds.), *Modeling and Using Context: 4th International and Interdisciplinary Conference CONTEXT 2003 Stanford, CA, USA, June 23-25, 2003 Proceedings* (pp. 164-177). Berlin: Springer.
- Guide to the ADL Gazetteer Content Standard, Version 3.2. (2004.). Santa Barbara, CA: Alexandria Digital Library Project. Retrieved July 11, 2007, from <http://www.alexandria.ucsb.edu/gazetteer/ContentStandard/version3.2/GCS3.2-guide.htm>
- Harpring, P., Beecroft, A., Johnson, R., & Ward, J. (Eds.). (2006). *Union List of Artist Names: Editorial Guidelines*. Los Angeles, CA: J.Paul Getty Trust.
- Hazelton, K. (Ed.). (2007). *EduPerson Object Class Specification: Internet2 Middleware Architecture Committee for Education, Directory Working Group (MACE-Dir)*. Retrieved July 11, 2007, from <http://www.nmi-edit.org/eduPerson/internet2-mace-dir-eduperson-200604.html>
- Heaney, M. (2000). *An Analytical Model of Collections and their Catalogues, Third Issue*. UK Office for Library and Information Networking. Retrieved August 24, 2007, from <http://www.ukoln.ac.uk/metadata/rslp/model/amcc-v31.pdf>

- Heckmann, D., & Krüger, A. (2003). A User Modeling Markup Language (UserML) for Ubiquitous Computing. In P. Brusilovsky, A. Corbett & F. de Rosis (Eds.), *User Modeling 2003: 9th International Conference, UM 2003 Johnstown, PA, USA, June 22-26, 2003 Proceedings* (Vol. 2702, pp. 393–397). Heidelberg: Springer.
- Heckmann, D., Schwartz, T., Brandherm, B., Schmitz, M., & von Wilamowitz-Moellendorff, M. (2005). GUMO: The General User Model Ontology. In L. Ardissono, P. Brna & A. Mitrovic (Eds.), *User Modeling 2005: 10th International Conference, UM 2005, Edinburgh, Scotland, July 24-29, 2005: Proceedings* (pp. 428-432). Berlin: Springer.
- Hedstrom, M. (2002). Archives, Memory, and Interfaces with the Past. *Archival Science*, 2(1-2), 21-43.
- Hedstrom, M. (1993). Descriptive Practices for Electronic Records: Deciding What is Essential and Imagining What is Possible. *Archivaria*, 36, 53-63.
- Hedstrom, M., & King, J.L. (2006). Epistemic Infrastructure in the Rise of the Knowledge Economy. In B. Kahin & D. Foray (Eds.), *Advancing Knowledge and the Knowledge Economy* (pp. 113-134). Cambridge, MA: MIT Press.
- Hedstrom, M., & Lee, C. A. (2002). Significant Properties of Digital Objects: Definitions, Applications, Implications. In *Proceedings of the DLM-Forum 2002, Barcelona, 6-8 May 2002: @ccess and preservation of electronic information: best practices and solutions* (pp. 218-227). Luxembourg: Office for Official Publications of the European Communities.
- Hedstrom, M. L., Lee, C. A., Olson, J. S., & Lampe, C. A. (2006). 'The old version flickers more': Digital Preservation from the User's Perspective. *American Archivist*, 69(1), 159-187.
- Hedstrom, M., & King, J. L. (2006). Epistemic Infrastructure in the Rise of the Knowledge Economy. In B. Kahin & D. Foray (Eds.), *Advancing Knowledge and the Knowledge Economy* (pp. 113-134). Cambridge, MA: MIT Press.
- Heidegger, M. (1996). *Being and time: a translation of Sien und Zeit* (J. Stambaugh, Trans.). Albany, NY: State University of New York Press.
- Hillier, B., & Hanson, J. (1984). *The social logic of space*. Cambridge: Cambridge University Press.
- Hoffmann, S. (2007). Processing Internet-derived Text—Creating a Corpus of Usenet Messages. *Literary and Linguistic Computing*, 22(2), 151-165.
- Hjørland, B., & Albrechtsen, H. (1995). Toward a New Horizon in Information Science: Domain-Analysis. *Journal of the American Society for Information Science*, 46(6), 400-425.
- Holdsworth, D., & Wheatley, P. (2004, April 13-16). 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. Retrieved July 12, 2007, from <http://romulus.gsfc.nasa.gov/msst/conf2004/Papers/MSST2004-03-Holdsworth-a.pdf>
- How long/large can my video be? Retrieved July 3, 2007, from <http://www.google.com/support/youtube/bin/answer.py?answer=55743&topic=10527>
- Hoyt, J., Daugherty, J., & Recordon, D. (2006). OpenID Simple Registration Extension 1.0. Retrieved July 16, 2007, from [http://openid.net/specs/openid-simple-registration-extension-1\\_0.html](http://openid.net/specs/openid-simple-registration-extension-1_0.html)
- Hu, Y., Li, H., Cao, Y., Meyerzon, D., & Zheng, Q. (2005). Automatic extraction of titles from general documents using machine learning. In *Proceedings of the 5th ACM/IEEE Joint Conference on Digital Libraries: Denver, CO, USA, June 7-11, 2005: digital libraries, cyberinfrastructure for research and education* (pp. 145-154). New York, NY: ACM Press.
- Hurley, C. (1995). Problems with Provenance. *Archives and Manuscripts*, 23(2), 234-259.
- Husserl, E. (1952). *Ideas: general introduction to pure phenomenology* (W. R. B. Gibson, Trans.). London: Allen & Unwin.

- ICA Committee on Descriptive Standards. (2004). *International Standard Archival Authority Record for Corporate Bodies, Persons and Families (ISAAR/CPF)*, Second Edition. Paris, France: International Council on Archives.
- IDEAlliance Publishing Requirements for Industry Standard Metadata: Guide to the PRISM Digital Image Management Metadata Encoding, Version 1.0. (2006). Alexandria, VA: International Digital Enterprise Alliance, Inc.
- IFLA Study Group on the Functional Requirements for Bibliographic Records. (1998). *Functional Requirements for Bibliographic Records: Final Report*. Saur, München.
- IFLA Working Group on Functional Requirements and Numbering of Authority Records. (2007). *Functional Requirements for Authority Data: A Conceptual Model (Second Review Draft)*. International Federation of Library Associations and Institutions. Retrieved July 9, 2007, from <http://www.ifla.org/VII/d4/Franar-ConceptualModel-2ndReview.pdf>
- Ingwersen, P., & Järvelin, K. (2005). *The Turn: Integration of Information Seeking and Retrieval in Context*. Dordrecht: Springer.
- An Introduction to the Global Trade Item Number (GTIN). (2006). Lawrenceville, NJ: GS1 US. Retrieved August 19, 2007, from [http://barcodes.gs1us.org/dnn\\_bcec/Documents/tabid/136/DMXModule/731/Command/Core\\_Download/Default.aspx?EntryId=59](http://barcodes.gs1us.org/dnn_bcec/Documents/tabid/136/DMXModule/731/Command/Core_Download/Default.aspx?EntryId=59)
- 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/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 9594-7. (2005). *Information technology -- Open Systems Interconnection -- The Directory: Selected object classes*. 5th ed.
- ISO 14721. (2003). *Space data and information transfer systems – Open Archival information system – Reference Model*.
- ISO/IEC 15944-1. (2002). *Information Technology – Business Agreement Semantic Descriptive Techniques – Part 1: Operational Aspects of Open-Edi for Implementation*.
- ISO 19107. (2003). *Geographic information – Spatial schema*.
- ISO 19108. (2002). *Geographic information – Temporal schema*.
- ISO/IEC 21000-2. (2005). *Information technology – Multimedia Framework (MPEG-21) – Part 2: Digital Item Declaration*. 2nd ed.
- ISO 21127. (2006). *Information and documentation – A reference ontology for the interchange of cultural heritage information*.
- ISO Project 27729. *International Standard Party Identifier (ISPI)*. Retrieved July 16, 2007, from <http://collectionsCanada.ca/iso/tc46sc9/27729.htm>
- Jakobson, R. (1987). *Language in literature*. K. Pomorska & S. Rudy (Eds.). Cambridge, MA: Belknap Press.
- James, W. (1890). *The principles of psychology*. New York, NY: H. Holt and Company.

- Kaptelinin, V., & Nardi, B. A. (2006). *Acting with technology: activity theory and interaction design*. Cambridge, MA: MIT Press.
- Karasti, H., Baker, K. S., & 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*, 15(4), 321-358.
- Kim, Y., & Ross, S. (2006). Genre Classification in Automated Ingest and Appraisal Metadata. In J. Gonzalo, C. Thanos, M. F. Verdejo & R. C. Carrasco (Eds.), *Research and advanced technology for digital libraries: 10th European conference, ECDL 2006, Alicante, Spain, September 17-22, 2006: proceedings* (pp. 63-74). Berlin: Springer.
- Kirsh, D. (2001). The Context of Work. *Human-Computer Interaction*, 16, 305-322.
- Lagoze, C., & Hunter, J. (2001). The ABC Ontology and Model. *Journal of Digital Information*, 2(2).
- Lagoze, C., Krafft, D.B., Payette, S., & Jesuroga, S. (2005). What Is a Digital Library Anymore Anyway? Beyond Search and Access in the NSDL. *D-Lib Magazine*, 11(11).
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: the embodied mind and its challenge to Western thought*. New York, NY: Basic Books.
- Latour, B. (1992). Where are the missing masses? The sociology of a few mundane artifacts. In W. Bijker & J. Law (Eds.), *Shaping Technology/Building Society: Studies in Sociotechnical Change* (pp. 225-258). Cambridge, MA: MIT Press.
- Lea, M., O'Shea, T., & Fung, P. (1995). Constructing the Networked Organization: Content and Context in the Development of Electronic Communications. *Organization Science*, 6(4), 462-478.
- Lee, C. A. (2005). *Defining Digital Preservation Work: A Case Study of the Development of the Reference Model for an Open Archival Information System*. Unpublished PhD Dissertation, University of Michigan, Ann Arbor, MI.
- Lee, C. A., Tibbo, H. R., & Schaefer, J. C. (2007). Defining what digital curators do and what they need to know: the DigCCurr project. In *Proceedings of the 2007 Conference on Digital Libraries* (pp. 49-50). New York, NY: ACM Press.
- Léger, L., Tijus, C., & Baccino, T. (2005). Effect of the Task, Visual and Semantic Context on Word Target Detection. In A. Dey, B. Kokinov, D. Leake & R. Turner (Eds.), *Modeling and Using Context: 5th International and Interdisciplinary Conference CONTEXT 2005, Paris, France, July 5-8, 2005, Proceedings* (Vol. 3554, pp. 278-291). Berlin, Germany: Springer.
- Leifer, E. M. (1991). *Actors as observers: a theory of skill in social relationships*. New York, NY: Garland.
- Lessig, L. (2006). *Code: Version 2.0*. New York, NY: Basic Books.
- Levy, D. M. (2001). *Scrolling Forward: Making Sense of Documents in the Digital Age*. New York, NY: Arcade.
- Liebowitz, S. J., & Margolis, S. E. (2000). Path Dependence. In B. Bouckaert & G. D. Geest (Eds.), *Encyclopedia of Law and Economics* (pp. 981-998). Cheltenham, UK: Edward Elgar.
- Light, M., & Hyry, T. (2002). Colophons and Annotations: New Directions for the Finding Aid. *American Archivist*, 65(2), 216-30.
- Lindblom, J., & Ziemke, T. (2003). Social Situatedness of Natural and Artificial Intelligence: Vygotsky and Beyond. *Adaptive Behavior*, 11(2), 79-96.
- Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project. (2002). Vancouver, Canada. Retrieved June 7, 2007, from <http://www.interpares.org/book/index.cfm>

- McCarthy, J. (1987) Generality in Artificial Intelligence. *Communications of the ACM*, 30(12), 1030-1035.
- McCarthy, J. (1993). Notes on formalizing contexts. In R. Bajcsy (Ed.), *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence* (pp. 555–560). San Mateo, CA: Morgan Kaufmann.
- McKay, A., & Yakel, E. (2006, September 13-20). The Archival Metrics Project and Beyond: Creating Tools to Share Knowledge of Archival Use and Users. Paper presented at the ICA-SUV Seminar, Reykjavik, Iceland. Retrieved July 16, 2007, from [http://www2.hi.is/Apps/WebObjects/Hi.woa/swdocument/1010097/Aprille\\_Yakel.pdf](http://www2.hi.is/Apps/WebObjects/Hi.woa/swdocument/1010097/Aprille_Yakel.pdf)
- Making and Optimizing Your Videos. YouTube Video Toolbox. Retrieved July 3, 2007, from [http://www.youtube.com/t/howto\\_makevideo](http://www.youtube.com/t/howto_makevideo)
- Malone, T. W. (1983). How do people organize their desks? Implications for the design of office information systems. *ACM Transactions on Information Systems*, 1(1), 99-112.
- Mani, A., & Sundaram, H. (2007). Modeling user context with applications to media retrieval. *Multimedia Systems*, 12(4-5), 339-353.
- Mani, I., Verhagen, M., Wellner, B., Lee, C. M., & Pustejovsky, J. (2006). Machine learning of temporal relations. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL*, July 17-18, 2006, Sydney, Australia (pp. 753-760). Morristown, NJ: Association for Computational Linguistics.
- MARC 21 Concise Format for Bibliographic Data. (2006). Library of Congress. Retrieved October 17, 2007, from <http://www.loc.gov/marc/bibliographic/ecbdhome.html>
- MARC Value List for Relators and Roles. (2003). Library of Congress. Retrieved July 17, 2007, from <http://www.loc.gov/marc/sourcecode/relator/relatorlist.html>
- The Maturing Modern. (1956, July 2). *Time*. Retrieved June 27, 2007, from <http://www.time.com/time/magazine/article/0,9171,891296,00.html>
- Matuszek, C., Cabral, J., Witbrock, M., & DeOliveira, J. (2006, March 27-29). An Introduction to the Syntax and Content of Cyc. Paper presented at the AAI 06 Spring Symposium: Formalizing and Compiling Background Knowledge and Its Applications to Knowledge Representation and Question Answering, Stanford, CA. Retrieved July 16, 2007, from [http://www.cyc.com/doc/white\\_papers/AAAI06SS-SyntaxAndContentOfCyc.pdf](http://www.cyc.com/doc/white_papers/AAAI06SS-SyntaxAndContentOfCyc.pdf)
- Merton, R. K. (1957). The Role-Set: Problems in Sociological Theory. *British Journal of Sociology*, 8(2), 106-120.
- Metadata Authority Description Schema (MADS). Library of Congress. Retrieved July 9, 2007, from <http://www.loc.gov/standards/mads/>
- Metadata Encoding and Transmission Standard: Primer and Reference Manual. (Draft). (2007). Washington, DC: Digital Library Federation. Retrieved June 7, 2007, from <http://www.loc.gov/standards/mets/METS%20Documentation%20draft%20070310p.pdf>
- Metadata Object Description Schema (MODS). Library of Congress. Retrieved July 10, 2007, from <http://www.loc.gov/standards/mods/>
- Metadata Standards Framework – Preservation Metadata (Revised). (2003). National Library of New Zealand. Retrieved July 16, 2007, from [http://www.natlib.govt.nz/files/4initiatives\\_metaschema\\_revised.pdf](http://www.natlib.govt.nz/files/4initiatives_metaschema_revised.pdf)
- Miller, C. R. (1994). Genre as Social Action. In A. Freedman & P. Medway (Eds.), *Genre and the New Rhetoric* (pp. 23-42). Bristol, PA: Taylor and Francis.

- Miller, D. P. (2000). Out from Under: Form/Genre Access in LCSH. *Cataloging and Classification Quarterly*, 29(1/2), 169-188.
- Minsky, M. L. (1986). *The society of mind*. New York, NY: Simon and Schuster.
- Mourelatos, A. P. D. (1978). Events, processes, and states. *Linguistics and Philosophy*, 2(3), 415-434.
- Naphade, M., Smith, J. R., Tesic, J., Chang, S.-F., Hsu, W., Kennedy, L., et al. (2006). Large-Scale Concept Ontology for Multimedia. *IEEE Multimedia*, 13(3), 86-91.
- National Information Standards Organization. (2006). *Data Dictionary—Technical Metadata for Digital Still Images*. (ANSI/NISO Z39.87. 2006). Bethesda, MD: National Information Standards Organization.
- Niles, I., & Pease, A. (2001). Towards a Standard Upper Ontology. In C. Welty & B. Smith (Eds.), *Formal Ontology in Information Systems: Collected Papers from the Second International Conference, October 17th-19th, 2001, the Cliff House on Bald Head Cliff Overlooking the Atlantic Ocean* (pp.-2-9). New York, NY: ACM Press.
- Nordland, L. (2004). The Concept of 'Secondary Provenance': Re-Interpreting Ac Ko Mok Ki's Map as Evolving Text. *Archivaria*, 58, 147-59.
- Norman, D. A. (1990). *The Design of Everyday Things*. New York: Doubleday.
- North American Industry Classification System. (2002). U.S. Census Bureau. Retrieved July 16, 2007, from <http://purl.access.gpo.gov/GPO/LPS26109>
- O\*NET Content Model. National O\*NET Consortium. Retrieved June 8, 2007, from <http://www.onetcenter.org/content.html>
- O\*NET-SOC Taxonomy. (2006). National O\*NET Consortium. Retrieved June 8, 2007, from <http://www.onetcenter.org/taxonomy.html>
- Open Video Project. Retrieved July 16, 2007, from <http://www.open-video.org/>
- Orlikowski, W. J. (2000). Using technology and constituting structures: A practice lens for studying technology in organizations. *Organization Science*, 11(4), 404-428.
- Oxford Dictionary of National Biography. Oxford University Press. Retrieved July 16, 2007, from <http://www.oxforddnb.com/>
- Oxford English Dictionary, Second Edition. (1989). Oxford, UK: Oxford University Press.
- Palmer, C., Zavalina, O., & Mustafoff, M. (2007). Trends in Metadata Practices: A Longitudinal Study of Collection Federation. In R. Larson, E. Rasmussen, S. Sugimoto & E. Toms (Eds.), *Proceedings of the 7th ACM/IEEE Joint Conference on Digital Libraries, Vancouver, British Columbia, Canada, June 18-23, 2007* (pp. 386-395). New York, NY: ACM Press.
- Pearce-Moses, R. (2005). Context. In *A Glossary of Archival and Records Terminology* (pp.90-91). Chicago, IL: Society of American Archivists.
- Penker, M., & Eriksson, H.-E. (2000). *Business Modeling With UML: Business Patterns at Work*. New York, NY: John Wiley & Sons.

- Petras, V., Larson, R. R., & Buckland, M. (2006). Time period directories: a metadata infrastructure for placing events in temporal and geographic context. In *Opening information horizons: 6th ACM/IEEE-CS Joint Conference on Digital Libraries: June 11-15, 2006, Chapel Hill, NC, USA: JCDL 2006* (pp. 151-160). New York, NY: ACM Press.
- Phelps, T.A., & Wilensky, R. (2000). Multivalent Documents: Anywhere, Anytime, Any Type, Every Way User-Improvable Digital Documents. *Communications of the ACM*, 43(6), 82-90.
- Powell, A., Heaney, M., & Dempsey, L. (2000). RSLP Collection Description. *D-Lib Magazine*, 6(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. Retrieved July 10, 2007, from <http://www.oclc.org/research/projects/pmwg/premis-final.pdf>
- Preyer, G., & Peter, G. (2005). *Contextualism in Philosophy: Knowledge, Meaning, and Truth*. New York, NY: Oxford University Press.
- Process. In Wikipedia. Retrieved July 16, 2007, from <http://en.wikipedia.org/wiki/Process>.
- Pustejovsky, J., Castaño, J., Ingria, R., Saurí, R., Gaizauskas, R., Setzer, & Katz, G. (2003). TimeML: Robust Specification of Event and Temporal Expressions in Text. In H. Bunt, I. van der Sluis & R. Morante (Eds.), *Proceedings of the IWCS-5 Fifth International Workshop on Computational Semantics*. Tilburg, Netherlands: Tilburg University, Computational Linguistics and AI Group.
- Recordkeeping Metadata Standard for Commonwealth Agencies, Version 1.0. (1999). Canberra: National Archives of Australia. Retrieved July 16, 2007, from <http://www.naa.gov.au/recordkeeping/control/rkms/contents.html>
- Report Concerning Space Data System Standards: Standard Formatted Data Units – A Tutorial, Green Book, Issue 1. (1992). Washington, DC: Consultative Committee for Space Data Systems.
- Robertson, B. G. (2006). Visualizing an historical semantic web with HEML. In *Proceedings of the 15th International Conference on World Wide Web*, Edinburgh, Scotland, May 23 - 26, 2006 (pp. 1051-1052). New York, NY: ACM Press.
- Robinson, C. (1997). Records Control and Disposal Using Functional Analysis. *Archives and Manuscripts*, 25(2), 288-303.
- Roe, K. (1992). Enhanced Authority Control: Is It Time? *Archivaria*, 35, 119-129.
- Roe, K. (2005). *Arranging & Describing Archives & Manuscripts*. Chicago, IL: Society of American Archivists.
- Rogers, E. M. (1995). *Diffusion of innovations* (4th ed.). New York, NY: Free Press.
- Ross, S. (2007). *Digital Preservation, Archival Science and Methodological Foundations for Digital Libraries*. European Conference on Digital Libraries.
- Rules for Archival Description. (2003). Ottawa, Ontario: Canadian Council of Archives.
- Rust, G., & Bide, M. (2000). The <indecs> metadata framework: Principles, model and data dictionary, WP1a-006-2.0. Retrieved July 16, 2007, from [http://web.archive.org/web/\\*/http://www.indecs.org/pdf/framework.pdf](http://web.archive.org/web/*/http://www.indecs.org/pdf/framework.pdf)
- Saurí, R., Knippen, R., Verhagen, M., & Pustejovsky, J. (2005). Evita: A Robust Event Recognizer For QA Systems. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing* (pp. 700-707). Morristown, NJ: Association for Computational Linguistics.
- Schilit, B. N., & Theimer, M. M. (1994). Disseminating active map information to mobile hosts. *IEEE Network*, 8(5), 22-32.
- Scriberras, A. (Ed.). (2006). *Lightweight Directory Access Protocol (LDAP): Schema for User Applications*. (RFC 4519).

- Sellen, A. J., & Harper, R. (2002). *The myth of the paperless office*. Cambridge, MA: MIT Press.
- Shah, P., Schneider, D., Matuszek, C., Kahlert, R. C., Aldag, B., Baxter, D., et al. (2006). Automated Population of Cyc: Extracting Information about Named-entities from the Web. In G. Sutcliffe & R. Goebel (Eds.), *Proceedings of the Nineteenth International Florida Artificial Intelligence Research Society Conference* (pp. 153-158). Menlo Park, CA: AAAI Press.
- Shanon, B. (1990). What Is Context? *Journal for the Theory of Social Behaviour*, 20(2), 157-66.
- Sharer, R. J., & Ashmore, W. (1987). *Archaeology: discovering our past*. Palo Alto, CA: Mayfield.
- Sibille, C. (2007, May 25). A New International Standard for Describing Functions: ICA-ISDF. Presented at The Standard ISDF: One More Step in Standardization, Ávila, Spain. Retrieved July 10, 2007, from [http://www.acal.es/portals/4/ACAL\\_Claire\\_Sibille\\_ICA-ISDF\\_Ingles.pdf](http://www.acal.es/portals/4/ACAL_Claire_Sibille_ICA-ISDF_Ingles.pdf)
- Smith, B. (2007). On Place and Space: The Ontology of the Eruv. In C. Kanzian & E. Runggaldier (Eds.), *Cultures: conflict - analysis - dialogue; proceedings of the 29. International Ludwig Wittgenstein Symposium, Kirchberg am Wechsel, Austria 2006* (pp. 403-416). Ontos: Frankfurt.
- Smith, B. C. (1996) *On the Origin of Objects*. Cambridge, MA: MIT Press.
- Smith, B., & Varzi, A. C. (2000). Fiat and Bona Fide Boundaries. *Philosophy and Phenomenological Research*, 60(2), 401-420.
- Smith, D. A. (2002). Detecting and Browsing Events in Unstructured Text. In M. Beaulieu, R. Baeza-Yates, S. H. Myaeng & K. Jarvelin (Eds.), *SIGIR 2002: proceedings of the Twenty-Fifth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, August 11-15, 2002, Tampere, Finland* (pp. 73-80). New York, NY: ACM Press.
- Smith, D. A., & Crane, G. (2001). Disambiguating Geographic Names in a Historical Digital Library? In P. Constantopoulos & I. Sølvyberg (Eds.), *Research and advanced technology for digital libraries 5th European conference, ECDL 2001, Darmstadt, Germany, September 4-9, 2001: proceedings* (pp. 127-136). Berlin: Springer.
- Smythe, C., Tansey, F., & Robson, R. (2001). *IMS Learner Information Packaging Information Model Specification, Version 1.0*. Lake Mary, FL: IMS Global Learning Consortium.
- Snow, D. R., Gahegan, M., Giles, C. L., Hirth, K. G., Milner, G. R., Mitra, P., et al. (2006). *Cybertools and Archeology. Science*, 311, 958-959.
- Society for American Archeology. (1996). *Archeology Terms*. Retrieved June 7, 2007, from <http://www.saa.org/publications/sampler/terms.html>
- Source Codes for Genre. Library of Congress. Retrieved July 16, 2007, from <http://www.loc.gov/marc/sourcecode/genre/genresource.html>
- Souza, C. d., Froehlich, J., & Dourish, P. (2005). Seeking the source: software source code as a social and technical artifact. In K. Schmidt, M. Pendergast, M. Ackerman & G. Mark (Eds.), *GROUP '05: proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work: November 6-9, 2005, Sanibel Island, Florida, USA* (pp. 197-206). New York, NY: Association for Computing Machinery.
- Sperberg-McQueen, C. M. & Burnard, L. (Eds.). (2002). *TEI P4: Guidelines for Electronic Text Encoding and Interchange. Text Encoding Initiative Consortium*. Retrieved July 11, 2007, from <http://www.tei-c.org/P4X/>

- Stalnaker, R. C. (1978). *Assertion. Syntax and Semantics*, 9, 315-322.
- Standard Occupational Classification System. (2000). Washington, DC: U.S. Department of Labor, Bureau of Labor Statistics. Retrieved June 8, 2007, from <http://stats.bls.gov/soc/>
- Suchman, L. A. (1987). *Plans and situated actions: the problem of human-machine communication*. Cambridge: Cambridge University Press.
- Suderman, J. (2001, June). *Context, Structure and Content: New Criteria for Appraising Electronic Records*. Paper presented at the Annual Conference of the Association of Canadian Archivists. Retrieved July 16, 2007, from [http://www.rbarray.com/suderman-wholepaper7\\_postscript011102.htm](http://www.rbarray.com/suderman-wholepaper7_postscript011102.htm)
- Suggested Upper Merged Ontology (SUMO). Retrieved June 14, 2007, from <http://www.ontologyportal.org/>
- Sweet, M., & Thomas, D. (2000). *Archives Described at Collection Level*. *D-Lib Magazine*, 6(9).
- Thibodeau, S. (1995). *Archival Context as Archival Authority Record: The ISAAR(CPF)*. *Archivaria*, 40, 75-85.
- Thiel, U., Brocks, H., Frommholz, I., Dirsch-Weigand, A., Keiper, J., Stein, A., & Neuhold, E.J.. (2004). *COLLATE – A collaboratory supporting research on historic European films*. *International Journal on Digital Libraries*, 4(1), 8-12.
- Thoma, G. R., Mao, S., & Misra, D. (2005). *Automated Metadata Extraction to Preserve the Digital Contents of Biomedical Collections*. In J. J. Villanueva (Ed.), *Proceedings of the 5th IASTED International Conference on Visualization, Imaging and Image Processing 2005* (pp. 214-219). Calgary, Canada: ACTA Press.
- Thomas, W., Gregory, A., Kuo, I.-L., Wackerow, J., & Nelson, C. (Eds.). (2007). *Data Documentation Initiative (DDI) Technical Specifications, Part I: Overview (Version 3.0) for Public Review: DDI Alliance*. Retrieved July 16, 2007, from <http://www.ddialliance.org/DDI/ddi3/overview.pdf>
- Tibbo, H.R. (1993). *Abstracting, Information Retrieval, and the Humanities*. Chicago, IL: American Library Association.
- Transportation Research Thesaurus. 2006. Transportation Research Board. <http://ntlsearch.bts.gov/tris/trt.do>
- Tyler, A., & Evans, V. (2003). *The Semantics of English Prepositions: Spatial Scenes, Embodied Meaning, and Cognition*. Cambridge, UK: Cambridge University Press.
- VRA Core 4.0. (2007). Visual Resources Association. Retrieved July 5, 2007, from <http://www.vraweb.org/projects/vracore4/index.html>
- Wallace, D. A. (1995). *Managing the Present: Metadata as Archival Description*. *Archivaria*, 39, 22-32.
- Wang, X., Demartini, T., Wragg, B., Paramasivam, M., & Barlas, C. (2005). *The MPEG-21 Rights Expression Language and Rights Data Dictionary*. *IEEE Transactions on Multimedia*, 7(3), 408-17.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: methods and applications*. Cambridge, UK: Cambridge University Press.
- Weber, M. B., & Favaro, S. (2007, April 18-20). *Beyond Dublin Core: Development of the Workflow Management System and Metadata Implementation at Rutgers, The State University of New Jersey*. Paper presented at DigCCurr2007: An International Symposium in Digital Curation, Chapel Hill, NC. Retrieved July 16, 2007, from [http://www.ils.unc.edu/digccurr2007/papers/weberFavaro\\_paper\\_4-1.pdf](http://www.ils.unc.edu/digccurr2007/papers/weberFavaro_paper_4-1.pdf)
- Weick, K. E. (1995). *Sensemaking in Organizations*. Thousand Oaks, CA: SAGE Publications.
- White, S. A. (2006). *Business Process Modeling Notation Specification*. Needham, MA: Object Management Group.

- Wiberley, S., & Jones, W. G. (1989). Patterns of information seeking in the humanities. *College & Research Libraries*, 50, 638-645.
- Wildemuth, B. M., Marchionini, G., Wilkens, T., Yang, M., Geisler, G., Fowler, B., et al. (2002). Alternative Surrogates for Video Objects in a Digital Library: Users' Perspectives on Their Relative Usability. In M. Agosti & C. Thanos (Eds.), *Research and advanced technology for digital libraries: 6th European conference, ECDL 2002, Rome, Italy, September 2002: proceedings* (pp. 493-507). Berlin: Springer.
- Wilson, A., & Clayphan, R. (2004). *Functional Requirements for Describing Agents, Draft 2*. Dublin Core Metadata Initiative - Agents Working Group. Retrieved July 15, 2007, from <http://dublincore.org/groups/agents/agentFRdraft2-2.html>
- Winner, L. (1986). Do Artifacts Have Politics? In L. Winner (Ed.), *The whale and the reactor: a search for limits in an age of high technology* (pp. 19-39). Chicago, IL: University of Chicago Press.
- Winograd, T., & Flores, F. (1986). *Understanding Computers and Cognition: A New Foundation for Design*. Norwood, NJ: Ablex.
- WordNet. Version 3.0. (2006). Princeton University Cognitive Science Laboratory. Retrieved June 13, 2007, from <http://wordnet.princeton.edu/>
- XML Process Definition Language, Version 1.15. (2005). Hingham, MA: Workflow Management Coalition.
- Yang, M., & Marchionini, G. (2005). Deciphering visual gist and its implications for video retrieval and interface design. In W. A. Kellogg, S. Zhai, C. Gale & G. C. van der Veer (Eds.), *CHI 2005: technology, safety, community: conference proceedings: Conference on Human Factors in Computing Systems: Portland, Oregon, USA, April 2-7* (pp. 1877-1880). New York, NY: Association for Computing Machinery.
- Yates, J., & Orlikowski, W. (1992). Genres of Organizational Communication: A Structural Approach to Studying Communication and Media. *Academy of Management Review*, 17(2), 299-326.
- Yi, K., Beheshti, J., Cole, C., Leide, J.E., & 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*, 57(9), 1208-20.
- Yu, B. 2006. *An Evaluation of Text Classification Methods for Literary Study*. Unpublished PhD Dissertation, University of Illinois, Urbana, IL.
- Zeilenga, K. D., ed. (2006). *COSINE LDAP/X.500 Schema*. (RFC 4524).
- Zimmerman, A. S. (2003). *Data sharing and secondary use of scientific data: Experiences of ecologists*. Unpublished PhD Dissertation, University of Michigan, Ann Arbor, MI.