Routledge Studies in Library and Information Science

1. Using the Engineering Literature
Edited by Bonnie A. Osif

2. Museum Informatics
People, Information, and Technology in Museums
Edited by Paul F. Marty and Katherine Burton Jones

Previous titles to appear in Routledge Studies in Library and Information Science include:

Using the Mathematics Literature
Edited by Kristine K. Fowler

Electronic Theses and Dissertations
A Sourcebook for Educators: Students, and Librarians
Edited by Edward A. Fox

Global Librarianship
Edited by Martin A. Kesselman

Using the Financial and Business Literature
Edited by Thomas Slavens

Using the Biological Literature
A Practical Guide
Edited by Diane Schmidt

Using the Agricultural, Environmental, and Food Literature
Edited by Barbara S. Hutchinson

Becoming a Digital Library
Edited by Susan J. Barnes

Guide to the Successful Thesis and Dissertation
A Handbook for Students and Faculty
Edited by James Mauch

Electronic Printing and Publishing
The Document Processing Revolution
Edited by Michael B. Spring

Library Information Technology and Networks
Edited by Charles Grosch

Museum Informatics

People, Information, and Technology in Museums

Paul F. Marty

and

Katherine Burton Jones

Routledge
Taylor & Francis Group
New York London

brought new expectations and new opportunities. As museum researchers and professionals continue to explore new ways of representing information about museum resources, they are radically changing the way museum professionals, visitors, and all users of those resources work with museum collections. When examining these changes, it is all too easy to be captivated by their novelty and potential. It is important to remember, however, that these changes are built upon a solid historical foundation of information representation in museums.

4 Representing Museum Knowledge

David Bearman

Archives and Museum Informatics

INTRODUCTION

Collectively museums hold the universe of all objects and ideas and all their relations. They have assumed responsibility for preserving what is now thought about them and what has been thought in the past, and for representing and interpreting that for the present. Because museums are collections, they have made even nature into a cultural artifact (Buckland, 1997). Thus, representing museum knowledge is potentially a task as comprehensive as the representation of all human knowledge.

The act of collecting has privileged those attributes of the object around which the collection is constructed and deprecates others, but decisions about representation should enable the object to be re-incorporated logically into many collections and contexts, including their original context, to support the work that those in and outside museums do with museum objects.

Museum knowledge representation has acquired additional requirements as a consequence of the computerization of much museum work and museum relations with visitors. Though computers were used to inventory museum collections from the 1960s on, computer representations of museum holdings evolved in sophistication from the mid-1980s to mid-1990s, as computing systems became increasingly capable of holding extensive museum data and data models were developed to support more and more museum work processes. Since the mid-1990s, the advent of the World Wide Web and networked computing has radically transformed the task as it was previously understood, in particular by redefining its audience, and thereby forced museums to rethink the purposes and ways they represent knowledge. This chapter proposes some guidelines for the present that can be gleaned from prior museum practice and other frameworks for representations. It illustrates how radically our concepts of what it is crucial to represent have changed over the past three decades, suggesting that today's view will be found lacking soon, but nevertheless attempting to guide current knowledge representation practices.

OBJECTIVES OF MUSEUM KNOWLEDGE REPRESENTATION

While museums do strive to collect every natural and human-made thing that has ever been and to interpret everything that has ever been thought about them, museum knowledge representation supports museum missions.

Museums make and record meanings. It has been argued that information management is, therefore, the central purpose of museums (Washburn, 1984). Certainly, the knowledge model of museums is much more robust than the actual accumulation of knowledge about any given object. The model can serve to highlight lacunae in our knowledge, directing research and documentation. Through research museums seek to document the material and ideational world as they, the original discoverers, and the creators of the objects and specimens they acquire, understood them. These are, of course, diverse and potentially conflicting, perspectives.

Museums seek to convey their understandings of their collections to scholars and lay people, experts and the naïve, adults and children. As such, they strive to articulate their knowledge in many different ways, to different depths and at different levels of sophistication. The museum must hold all these representations at one time if they are to be presented to their desired audiences.

Museums seek to preserve their holdings and knowledge, not just over time, but also from one product or process to the next. The information that is recorded should, therefore, serve the purposes of each of those museum activities that need to use it, requiring that it be encoded in a way that makes it efficiently suited to the purposes for which it will be used.

These three goals—making, conveying and preserving meanings—will be explored in this essay as we delineate the domain of museum knowledge representation and suggest guidelines that might govern how we can best describe, explain and control museum objects.

THE CHALLENGE OF KNOWING

To understand what it means to represent knowledge in the museum, we need the humility to appreciate that our knowledge of the world is socially constructed. Museums strive to represent what they know, but what they know was conveyed to them by someone who first made or discovered the object, collected or analyzed it, or acquired it for the museum or managed it in the collection. When we say what something "is," it cannot be said to be "true" though it may be preferable for particular purposes or "correct" from a specific perspective. To the explorer who first collected the artifacts, the people he had encountered were Eskimo; to themselves they were "Inuit," meaning, "the people." One informant may be as certain that an object is a jaguar as another is that it is the spirit of his grandfather. The fact of the social construction of all knowledge means that in museums, all knowledge should be sourced. Yet this is one of the least observed requirements of knowledge representation systems in museums, and results in one of the most criticized aspects of museum interpretation, its adoption of an authoritative, unsourced voice (Walsh, 1997).

We must not lose sight of the fact that objects in museums have been collected from some natural or cultural context in which they originated or have been used. The museum is a storehouse of things that were consciously gathered and placed in the context of other things also gathered. We need to be aware that the representations we have were made for a purpose. Thus we might know that an object was acquired at dusk, or in the spring, or on the birthday of the collector, or in 1842, all reflecting quite different purposes and assumptions about what is significant to the object, to the act of collecting and to the different social constructs in which we make sense of such things. Any museum object has several stories to tell: the story of having been collected might be thought of as their stories as told by their original collectors, while other stories are those told by subsequent curators or researchers. Yet too frequently our abstract frameworks for representing what we know assume a singular point of view about what is worth recording and how. For example, if the day, month and year that an object was collected are the only form of "time of collection" supported, we are deprecating other perspectives that in other contexts might be more relevant.

What we know is further qualified by why we know it. We might know of an object that it was given in tribute because what we know about it was recorded by the recipient; what might have been said by the "gift giver"? In addition to lacking information from all possible sources, we are always at risk of substituting our cultural perception for that of others. We might conclude that an item was acquired by theft, for having been found in a "hoard" of objects seemingly pillaged, but it might upon further study be a kind of bank, to which voluntary deposits were made. We might "know" that a stone to which magical properties were associated by the peoples who owned it is a strong lodestone, but they did not "know" that and recording our knowledge does not alter theirs. By chemical analysis we might know that a pigment on a famous masterpiece which "hung on the mantel of a major local landowner since the 16th century" was not invented until the 18th century, yet this "fact" would not change the role this object's place in family history at all. Collections of facts, like collections of the objects to which they relate, are built up over time and have a life of their own; when we represent both a thing and the knowledge of a thing, we must be prepared for divergence between the two.

What we represent and how we encode it should be faithful to the evidence that we have. Thus if the testimony accompanying an object dates it from "the third year of the Depression," recording it as 1932 loses significant information given us by the informant and privileges the curator's interpretation of the dating of the Depression. Not recording 1932 makes it hard to correlate the object to other objects whose dates of creation are known years, so we may need to keep that representation too. In documenting the object in the museum, we are recording the history of its interactions with other objects, and with people, places, events and actions. A "hard" fact, like the date an object was first created or used, may involve documenting the testimony of the donor, the interpretation of the curator and the evidence of a scientific analytic tool, in each case recording the "date" and who claimed what, when. In the museum we need to record many forms of date as they convey different information, reflecting their different relationship to the source of the data, even though they may not refer to different years.

THE CHALLENGE OF RECORDING

Some ways of representing information will be better suited than others to particular subsequent uses. When the facts about objects are recorded, we tend to record them from the perspective of the activity in which the act of recording takes place. For museums, these activities might be collection, acquisition, conservation, exhibition, interpretation, or research for example. In each case, there is a time and place in which the documentation occurs; there is an observer who is recording, possibly with tools to assist, and with ideas involved. All these facts about the context of the activity and its bearing on the object need to become part of the museums' knowledge. Later there will be reasons, of accountability or of curiosity, to recall these facts.

Time, for instance, is an omnipresent aspect of the knowledge we have of objects, since the museum gathers the past. The periods measured in museums range from nanoseconds to astronomical units of time, and the schemes employed include all those ever used by man. For instance, if we are concerned for conservation to determine if an object in our storerooms was subjected to a specific danger, we will want to know the date and time of day it was moved there. If we are dating a geological specimen in our collection for researchers, a date expression of $3.2^{-6} \pm 5^{-3}$ will better support calculation and reflect what our carbon dating instruments actually told us. The reading from our instruments has the advantage that it will never change, but the name of the geological era, like any interpretation, could be reconsidered and the object we own could "move" to a different period. Nevertheless, if we are going to prepare a label for an exhibition, the geological period will probably be a better representation since it can be used with appropriate illustrative wall labels to locate the object in geologic times and most of the public finds mathematical expressions off-putting. The "correct" scheme with which to represent time in each *display* is the way that will be useful to the tasks at hand, but at the same time we need to *preserve* the primary form and source of data intact.

But selecting a representation scheme is not the end of the task. We must still choose how best to encode time. In our example of the time of storage, we'll no doubt want to choose "ANSI-time" or YYYYMMDD (which can be extended with HHMMSS etc.) over the prose expression June 15, 1999 because it will be easiest to manipulate in our computing system. But if we are recording the time that an artifact in our collections was created or used, we might be confronted with a more complex decision. Rather than using ANSI time to record a date range from the Gregorian calendar with which we are familiar, we might wish to represent the time of creation in terms that the creator recorded them. Prior to the mid-16th century in the West, this would have been the Julian calendar, but elsewhere it might have been based on the Zoroastrian, Jewish, Chinese, or Mayan calendars, each of which employs quite different units of time and "begins" at different dates. Or it might be a more accurate reflection of our knowledge to use culturally referential periods of time such as "the Depression" or "the Ming Dynasty" or "the time of Abraham" if we are reasoning about the time of creation and use based on other observations that we associate with these periods rather than based on recorded dates.

One more example should suffice to make the point. In our administrative recordkeeping within the museum, personal names of employees or donors for example might be represented with discrete elements for Titles, First Name, Middle Name, Last Name, but could be displayed in inverted order as "Last, First, Middle, Title" or in direct order "Title, First, Middle, Last." Each conveys the same information, but the latter is preferred for sorting lists by last name, while the former is the preferred form of address; both are enabled by representing the units discretely. But for other purposes the museum needs to know that this form of representation and encoding is very culturally specific. To avoid that, the museum could keep the same representation, but label the elements "surname" or "patronymic" and "given names" or "personal names." But if the museum needs to manage names from many cultures and periods it needs to consider whether "toponymics" might be importantly different from other "last names" and realize that names of individuals, as a category, is probably too narrow a construct to represent the makers and creators of artifacts in a universal collection. In our society corporate entities are often creators, and in other cultures and times the creation of cultural artifacts would not be attributed to individuals at all, but to villages or totemic groups or other cultural units.

The choice of how to represent information, whether it is a personal name, time or any other attribute, should not be arbitrary. There are criteria by which we can distinguish useful from less useful representations.

How we encode what we know should be useful for the purposes that we have, and as useful as possible to purposes we may have, but do not yet know. Always the information should be expressed in a way that is explicit and can be used as flexibly as possible in the tasks that we know it will be involved in, within the museum and by others. Thus, even if the museum "has always used cm height x width x depth in its measurements," a measurement that explicitly enumerates the measurement units and dimensions measured—14 cm h x 3 cm w x 2 cm d —is preferable to recording 14 x 3 x 2 as NASA recently learned at the expense of losing a multi-billion dollar planetary mission (Lloyd, 1999). In addition, the structure of recording the quantity, the unit and the dimension permits us to add our measurement of the base as 4 cm diameter and the mouth as 8 cm diameter, at some point in the future as our knowledge of the object develops.

Increasingly as we consider representation of knowledge we need to remember that we can record facts and opinions about objects not only in words or in measurements and scientific analyses, but also in images, in sound and in multimedia. When we do this today, we will encode all these diverse representations in binary form. Thus we have at least four levels of decisions that we must make about documentation—what features to document, what about those features to represent, how to express what we want to record, and how to digitally encode it. Each of these decisions will impact on what the representation that we make can be made to "do." Metadata about how these representations were made and how they are encoded is essential to their subsequent reuse.

Finally, effective encoding should aim to be of continuing value. This means we need to consider not just the purposes for which it was originally recorded, but future purposes as well. Historically museums have been intellectually profligate. They have expended vast quantities of the time of highly trained staff creating numerous representations of small parts of their collections and throwing each subsequent representation away when its work was done. Consider the exhibit, in which each of a small number of objects is re-researched, re-described and re-interpreted only to have these expensively constructed knowledge components discarded along with an equally expensive construct that holds them at the end of a brief public run. Yet to use these representations, we need to not only document who made them, why, and what purposes they served, but to do so in a manner that will be consistent with representations we have kept from other sources so that we can efficiently decide on their re-use. This places a huge burden on our choice of encodings.

A BRIEF HISTORY OF MUSEUM KNOWLEDGE REPRESENTATION

Object-Centered Data Models

The creation of museums and the keeping of records about collections of artifacts and specimens are historically intertwined. Field notebooks of naturalists, ship logs from expeditions of maritime exploration, accounts by the bookkeepers of princely hoardings, country house inventories, and registries of wills supplement museum ledgers recording acquisitions over the first few centuries of museum history. In each case, these are records of transactions, organized around the activity of recording, in chronological sequence.

Card catalogs and vertical files augmented these transaction records during the early 20th century, providing some access by object, though the principal means of accessing individual objects in museums remained the organization of their storage and the records of their acquisition. Card systems tended to reflect the idiosyncratic interests of curators and were often abandoned after their retirement, or replaced by new systems with different orientations. Projects external to museums, such as the Princeton University Index of Christian Art or the ICONCLASS classification system invented by Henri van del Waal at the University of Utrecht, evolved "encodings" for recording and searching museum documentation, but these were secondary resources. Until the arrival of computers, museum documentation was not generally organized around object-centered knowledge representation models designed to expedite retrieval and use based on properties of the thing, rather than on the history of its relation to the museum. The history of computers in museums is addressed elsewhere (Jones-Garmil, 1997) and in Chapter 2 of this volume; here we will look only at aspects crucial to understanding museum knowledge representation.

The registrars and computer scientists who formed the Museum Data Bank Committee (subsequently MCN) viewed the computer as a tool to create a catalog of the collection, just as librarians at the same time were imagining it as a means of compiling bibliographic catalogs. The systems they developed were collection inventory files processed by mainframe computers fed initially by punch card input. In the 1970s, the literature on museum computing consisted in large part of the data dictionaries of systems that were designed to serve as centralized repositories of the metadata from many museums, and very basic presentations on the nature of museum computing systems by their developers (Bergengren, 1979; Chenhall, 1975, 1978; Gautier, 1979; Porter, 1979; Roller, 1976).

Prior to the mid-1980s, data dictionaries of implemented museum systems which circulated in loose-leaf binders, and assembly language object code distributed by the agencies promoting these systems, were the primary sources of any information there was about how museum databases were or ought to be constructed. "Published" dictionaries that were influential included those of the Smithsonian Institution (SELGEM system), the Metropolitan Museum of Art and Museum Databank Committee (GRIPHOS system), and the Canadian Heritage Information Network (CHIN Data Dictionaries, Sciences and Social Science/Humanities). The secondary works focused on the nature of computers and project management and did not treat knowledge representation (Chenhall, 1975; Orna & Pettitt, 1980, 1998; Van Someren Cok, 1981; Williams, 1987).

Although intended to be normative, these data dictionaries reflected local representation and encoding practices. All represented museum objects as records of very limited length in flat files created for mainframe computers. None of them reflected explicitly on alternative choices in knowledge representation or discussed the limitations of the approaches they had taken. The data values were all presumed to be objective. The data were unattributed, and discussions of it were utterly non-reflective about their methods of recording. Limited lengths of fields encouraged abbreviations and look-up tables, and made prose impossible. Pre-relational systems restricted description of other entities to whatever characteristics could be attributed to the collection item.

In the mid-1980s, the author introduced formal data modeling and relational databases to the Smithsonian Institution, and elsewhere in the world these new computing methods, the spread of micro-computers and cathode ray displays, and the falling prices of storage, had a dramatic impact enabling choices in representation of museum data and its active use in support of museum missions. From the mid-1980s on, the Museum Documentation Association in the United Kingdom and Archives & Museum Informatics in the United States moved beyond the dissemination of data models to the promotion of museum standards for recording and handling data content and a focus on the functionality that should be associated with different types of museum information systems (Bearman, 1987, 1990a, 1990b; Light, Roberts, & Stewart, 1986; Roberts, 1985, 1988, 1993).

The rise of commercial software which spread with the mini-computer (and later the micro-computer) was an alternative to custom developed mainframe applications. But to exploit this opportunity, museum professionals needed to be able to compare different systems, and assess them based on criteria including their interoperability with other systems that were imagined (in imitation of the emerging "integrated library systems"), as functional modules of a to-be-realized integrated museum system. These comparisons were published in bi-annual volumes of the Directory of Software for Archives and Museums (Bearman & Cox, 1990; Bearman & Wright, 1992, 1994), accompanied by essays by the editor highlighting important knowledge representation and functionality developments.

A new generation of more critical analyses arrived in the mid- to late-1980s (Abell-Seddon, 1988, 1989; Bearman, 1987; Chenhall & Vance, 1988; Roberts, 1985). All these addressed knowledge representation as an issue in its own right, rather than as simply a question of encoding, storage efficiency and data preparation. All took the then relatively new perspective of the relational database to explore relations between entities beyond the collection object in itself and suggested departures from flat files. Each distinguished between issues relating to syntax of representations, encoding rules and the semantics of representation.

By the 1990s it was possible to refer to a number of guidelines for integrated museum systems functionality, and a sophisticated, community developed, relational data model from CIDOC, the International Council on Museums Committee on Documentation (International Council of Museums). The CIDOC relational model, the culmination of years of consensus building, remains the best single statement of the relationship between structured elements of information about an object in the context of museum practice.

Efforts by humanists and scientists over many years have yielded insight into the many complex relations of scientific specimens (Allkin, White, & Winfield, 1992; Chavan & Krishnan, 2003; Graham, Ferrier, Huettman, Moritz, & Peterson, 2004). A multi-year funded effort to map the knowledge structure of art and architecture artifacts yielded the Categories for Description of Works of Art (Baca & Harpring, 1996; Trant, 1993), an important conceptual mapping that some have unfortunately tried to implement as a data structure rather using the reasoning that went into it to better understand the complexity of concepts in the field.

Process-Centered Data Models

In the early 1990s, the Museum Documentation Association under Andrew Roberts decided that the shift to object-centered recording in museums was creating a new documentation requirement, but not taking advantage of ongoing documentation activity which took place throughout the museum in all of its processes. They began to map the relationship between data about objects and the events within the museum that gave rise to that data. In 1994 the MDA issued SPECTRUM, a data standard organized around common procedures within museums and designed to re-integrate object documentation and museum workflow. The standard has undergone revision since (Museum Documentation Association, 2005), but remains the best single source of information on the way in which museum data is employed in museum processes, and indeed, of how museums actually work.

By the mid-1990s, pressure was building from other quarters in technology that would make the comprehensive statements of purely object-centered data relationships obsolete. First, computers were increasingly being used to store unstructured text, still images, and, ultimately, multimedia, not all of which represented the museum collection item, as reflected in the Proceedings of the ICHIM conference from 1991 to the present (Bearman, 1991, 1995c; Bearman & Garzotto, 2001; Bearman & Trant, 1997, 1999, 2003, 2004, 2005; Lees, 1993). Secondly, computers were increasingly used to communicate over networks, ultimately the Internet, which led to a growing interest in standards for data interchange and interoperability (Bearman, 1992b, 1995c; Bearman & Perkins, 1993), and also to a growing audience of non-specialist users, culminating in the general public. Thirdly, alternatives to the relational data model, driven in part by the need to exploit both these developments, and in part by a vision of object life-history that did not privilege the museum context over other periods in the objects' life, led to a proliferation of interest in object-oriented models and methods (Bearman, 1992a; Bearman & Vulpe, 1985; Research Libraries Group, 1994; Vulpe,

1986) and in semantic linking models (Beynondavies, Tudhope, Taylor, & Jones, 1994).

In 1994, these trends came together in a profoundly new implementation—the World Wide Web—which within a couple of years fundamentally transformed the methods, the audiences, and finally many of the purposes of museum knowledge representation. To an extent that has still not been fully understood, it brought museum representations into the same arena as those of other cultural institutions, where the exhibition and interpretation traditions of museums were highly successful paradigms. And with its popularity, and the spread of inexpensive networked computers, technology assisted workflow related computing requirements began to influence knowledge structuring practices in museums (Carliner, 2003; Marty, 1999).

In fact, over the past decade, while there has been much attention devoted to how museums can use the Web for a wide range of outreach purposes, as documented in the annual proceedings of the Museums and the Web conferences (see http://www.archimuse.com/conferences/mw.html/), there has been little explicit attention there or elsewhere to the implications of these developments for museum knowledge representation. In the remainder of this chapter, we will look at what was learned about museum knowledge representation from systems prior to the mid-1990s, examine the impact of the Web and the issues it presents for museum knowledge representation, and then hypothesize about requirements for future systems development and implementation.

LESSONS FROM PRE-1995 MUSEUM SYSTEMS

Formally Declared Data Models with Maximally Disaggregated Data

The relational model was built on a formal method of data normalization. Fully normalized data (fifth normal form) is maximally disaggregated, and while inefficient for any specific purpose, less normalized representations can be derived from it. For the first time, the choices of knowledge representations made in any given implementation could be explained with concrete reference to a logically derivable form of the data.

By naming data elements consistently, using entity-process-property-encoding conventions, it became possible to begin to map data within, and ultimately across, systems. Maker-birth-date and museum-acquisition-date or object-collection-place-geopolitical name and maker-death-place-geopolitical name, are two pairs of data values that will be expressed the same ways. This meant that we could define common routines to manipulate them and common indexes to search them.

By disaggregating, and by using formal decomposition methods and naming conventions, we were able to discover and then exploit common usages and usages that while different could be formally translated to be the same. Thus we were able to relate object-creation-geopolitical era to object-creation-date and object-creation-place and recognize that Ottoman Empire was both a time and a place, or rather is a place which has different boundaries at different times.

Experience in building museum systems had begun to convince some of its practitioners of the potential value of standards, in particular knowledge representation standards, which would govern how particular pieces of information, if present, would be expressed. The goal (though the term was not yet in use) was interoperability and exchangeability of data.

Multiple, Independent Authorities

Museum practitioners looked to libraries, which had been down the path quite successfully by the mid-1980s, for guidance in how best to represent knowledge for computer applications. In the context of 1980s information retrieval systems, which depended on pre-coordinated indices, libraries had demonstrated the benefits of "authority control" in searching for "known items."

Since then, museums have been on a quest for authority control, seeking the same benefits while neglecting the overwhelming differences in their holdings and of searching in the museum context. The results have been disappointing because of the significant difference between the presentist and retrieval orientation of libraries and the historical and contextualizing orientation of museums.

The most important difference between libraries and museums was so obvious that it was typically overlooked. Museum artifacts and specimens lacked "title pages" from which descriptive catalogers could "transcribe" the computer record. This difference in practice between transcription and attribution remains poorly understood in most comparisons of library, archives and museum documentation practices. Because publications almost always have known authors, titles and dates, searching in library catalogs is designed to retrieve known items. Artifacts and specimens almost always lack all these recorded metadata, so "known item" searching is quite atypical in the museum context.

Library catalogs, therefore, maximize the effectiveness of searching by collocating all items associated with a particular person, organization, or subject, by substituting "authorized" or "preferred" terms for data values that might otherwise be in the descriptive records. Hence, when persons were associated with artifacts or specimens in museum records, as their makers, designers, discoverers, owners, etc., they were frequently not in the Library of Congress Name Authority files. Often, of course, these people had not authored books, or they had under an assumed, literary name, which the Library of Congress "preferred," but even when the person was known to exist in a library authority, variations on names were considered "not preferred" and would never be used in the library setting. In contrast, museum practice would have dictated that all names by which someone had ever been known ought to be retained as each would be found in some documentation and reflected the point of view of the person using that form of the name. In their documentation, museums would normally use the name by which the person was known at the time and in the place that the attributed relation occurred. Similarly museums found that corporate creators of material culture tended not to be in the LC Corporate Names list because the names used by librarians reflected changes of ownership and not the name at the time of creation. Finally, name authorities were of little help in attributing works in the museum that had been made by anonymous individuals reflecting the methods and techniques known to their social groups. This applied to a large percentage of material objects, because attribution to individuals, rather than to groups, is a relatively modern and almost exclusively Western phenomenon.

An effort to develop a museum specific name authority for the arts, with more open ideas about the multiplicity of names an individual might have used during their lifetime or been referred to after their death, was undertaken by the Getty Trust (Bower, 1993; Siegfried & Bernstein, 1991), but it did not fully escape the limitations of prior practice. The practice of employing authorized or preferred names to reduce the particularism of the references to individuals within museum records erased the historical context of the associations between people and artifacts, and violated much museum practice and the architectures of searching through authorities to the object record that would have enabled the benefits of collocation without supplanting context, were insufficiently developed.

Similar difficulties bedeviled the use of other library authorities. Nonliterary cultural works tend not to have "titles," and when they do these are likely to be multiple and not well known. Artifacts do have common names, and the names of everyday things were compiled into a valuable "nomenclature" by Robert Chenhall in 1978. This nomenclature had numerous properties of a thesaurus and could be used in part to locate items of similar types. A more rigorous thesaurus, with a more focused topic, was developed by the Art and Architecture Thesaurus projects from 1979 to 1990, but it limited its terminology to that which had warrant in the published literature rather than drawing more widely from the realm of museum curatorial and scholarly terminology within the realm of practice (Petersen, 1990).

Subject terminology from the Library of Congress subject headings proved to be difficult to apply meaningfully. The Library of Congress Prints and Photographs Division Thesaurus for Graphics Materials incorporated "topical" subject terms that were more appropriate, but could not overcome the fact that museum specimens and artifacts do not "have" subjects (Parker, 1987).

Thus, before the advent of the Web, it had become evident that the library community's approach to vocabulary control was inappropriate for

museums. Whereas libraries decided that certain terms would best represent specific ideas to meet the information retrieval needs of their users today, the museum had three additional problems to contend with. First, the historical terms actually given to an object or concept over time were important properties of that object from the perspective of scholarly interpretation; more so than the information retrieval requirements of using today's language (Bearman, 1995a). Secondly, scholars almost by definition, do not agree on what constitutes reality; indeed one of the most important things to preserve in museum documentation was the on-going debate between scholars, and therefore multiple independent truths (Bearman, 1988; Bearman & Szary, 1986; Doerr, 1997). Finally, the perspectives of various users were all equally legitimate and quite different; the terminology that quite accurately describes a thing for one kind of user is not appropriate for another (Sledge, 1995).

Museums found themselves redesigning the purposes of authority files. Exemplifying a paradox of humanities computing, that it valued nuanced prose and a distinctive voice with which to address each clientele, museums resisted using authority control to engineer commonality. Instead, museums recorded situationally correct terminology and located it where and when it had been assigned, as well as by whom it was used and how it evolved over time. The knowledge representation needed to support this did not use controlled vocabularies to limit the terminology assigned to an object in the collection, but rather to expand it. By representing the range of what was known about people, places, actions and ideas without assuming the privileged position of the museum object, museums could architect search systems that positioned thesauri between users and the database so as to expand their search language (Bearman & Trant, 1998; Sledge, 1995). This meant that museums needed to learn the limits of term expansion both up and down thesaural hierarchies so that if a user searched for wood occasional tables they would find birch end tables and maple coffee tables (Bearman & Peterson, 1991).

Dozens of specialized vocabularies were developed from particular realms of curatorial practice at this time, but they in turn lacked both "literary warrant" and use outside their domain. Some were complex classification systems that could be applied by a trained observer (such as ICONCLASS) while others were ontologies specific to a narrow domain (such as the Railway Thesaurus). Discussions of these in the 1980s led nowhere, as there was no mechanical way to integrate them into a universal ontology nor any politically or intellectually acceptable way to give one precedence over another (Roberts, 1990). What the museum community discovered, long prior to the advent of RDF and the Semantic Web, was that it was necessary to employ multiple domain vocabularies since they served their communities uniquely, and that these could not be integrated in some universal ontology (Bearman, 1994). At the same time, a degree of integration between vocabularies could be supported by explicit sourcing of values.

Self-Consciously Universalistic Data Typing

By the mid-1990s, some museum knowledge representation standardization was gaining ground. Although commercial software developers had resisted interchange standards (Bearman & Perkins, 1993), fearing loss of market if museums could easily move data to other systems, they were adopting universalistic data typing as a common approach to defining complex data types.

Commercial applications developers had numerous clients for their software. Since they wanted both to make as few changes to their code as possible and yet name data in the ways their clients viewed it, they needed to be able to define their data "under the hood" in ways that could allow different museums, and indeed different curators within the same museum, to label and record it differently while maintaining a common knowledge architecture. They found that by supporting complex data types that were explicit about the intellectual system being represented, the tools and techniques of measurements, and the degrees of certainty of the recorded facts, and disaggregated the components of the attributes as fully as possible, they could support multiple world views.

Application systems in general use would support metric or Imperial measurements using two different hard coded routines, but museum software designers learned that museum application would want to record linear distance measurements using systems from other cultures and eras—such as the Greek (stadions and plethora) or Chinese (chi and li). In addition, some museums will want to record vast distances in astronomical units and cellular and atomic distances in angstroms. The range of possible measurement units required the disaggregation of a measured quantity, the unit, the system of measure, a dimension, degrees of certainty, and measurement methods (tools and techniques); in other words, the museum software developers learned to adopt self-consciously universalistic data typing practices.

Earlier we discussed expressions of time and periods of time. Self-consciously universalistic data typing required explicit recording of the dating system, period/era common name, early date year, early date month, early date day, early date expression of time, late date year, late date month, late date day, late day time, degree of certainty, instrument, method, when time was expressed numerically. It also required that other concepts of periods of time—such as political time-periods (World War II, the Ming Dynasty), cultural time-periods (the Victorian era, the Reformation), geological time-periods (the Pleistocene), personal time-periods (adulthood, pre-pubescence), each requiring different systems of measurement and different data structures for expressing values, be accommodated in the knowledge model.

Whereas place, or location of origin, in library descriptive cataloging means place of publication, and takes the values of geopolitical place names for a city and country, the location of origin for objects in museums can

be geo-morphological, geo-cultural or geo-linguistic, and geo-religious locations as well as geo-biological/botanical regions or even extra-terrestrial (oceanic and outer-space) locations.

In sum, in the museum, the "properties" of things are consequences of acts of knowledge declaration, and the tactic for making such declarations work over time is to adopt a self-consciously universalistic approach to data typing. In formal structures to represent time, space, events, and physical descriptions systems of representation are always explicitly declared. Individual data values—"x"—are replaced by tables that can qualify the data value by answering: "x" by what calendar? By what projection scheme? By what measuring instrument? And, by whom, where, when?

CONTEMPORARY MUSEUM KNOWLEDGE REPRESENTATION ISSUES

From 1985 to 1994, we saw the rise and spread of the personal computer which changed the character of office work worldwide and penetrated the household to some extent. But its effect on everyday life was trivial as compared to the influence of the World Wide Web and telecommunications-based computing since 1995. Indeed society as a whole, worldwide, has been significantly shaped in the decade since 1995 by the evolving nature of computing. The reasons are probably very simple—computing was awaiting the full integration of multimedia before it could be a populist medium.

Since the invention of the World Wide Web and http, we've seen a huge number of specific innovations in computing that have shown potential for museum applications. But the fact that vastly growing numbers of people are connected to the Internet every year and that they are spending vastly more time searching for things of interest to them (Fallows, 2005), explains why museums are trying to make themselves known to this audience and to compete for its attention rather than any particular technical synergy.

The success that museums have registered in increasing the size and variety of audiences visiting them on the Web has in turn promoted interests within the museum in further extending access. Most museums have taken accessibility of their Web sites quite seriously and are implementing W3C standards to further extend audiences. Active broad- and narrowcasting using RSS newsfeeds, blogs, and Webcasts are becoming more common. A few museums have begun to invest in embedded computing and smart buildings and are using these to individuate the information provided to visitors and to support a range of visitor information gathering studies. Each of these has further implications for knowledge representation, as do the creation of collaboration environments, intranets, and multi-platform, handheld consumer devices.

Re-Usable Metadata

The availability of a wide range of communications outlets (content platforms) all fed from digital data representations has increased the possible payback from making re-usable data. At the most basic level, museums are finding reasons to adopt XML as their standard for representation of text. As a practical matter, this means that textual components have explicit data types (albeit labels and behaviors are still based on local schemas, but at least they are identified) and that metadata can be directly displayed on the Web.

Arguments to adopt data standards for more complex objects are being considered by museum management. Object-oriented multimedia and calculated visualization methods including 3-D virtual reality are additional trends that require new knowledge representation. Location-aware data, whether to take advantage of mobile customers or GIS displays, are particularly demanding in this respect. As the audiences we reach grow and change, customer relationship management becomes important, which requires new metadata about new entities.

In sum, though neither multimedia nor telecommunications were new to museum computing, the advent of the World Wide Web, which marked their unification and extreme popularization, effectively transformed museum computing and knowledge representation.

Metadata for Management

The Museum and Its Programs

The museum Web site is a source of information about the museum and everything that it does. As a consequence, the data published to the Web from museum computing systems now has to include data about the museum itself, not just its collections, but its hours, its staff, and its services. This means not only a massive expansion of the quantity of data, but the involvement of virtually every department in publishing that data and keeping it up to date (Booth, 1998). Ultimately this means the scope of museum metadata expands to include process and responsibility, as well as audience appropriateness, language, and accessibility metadata. And it implies a need to keep track of versions of documents updated to the Web site at different times and by different departments and individuals.

The Museum and Its Societal Obligations

The museum is a legal institution and occupies a respected place in society which subjects it to various legal and ethical obligations with implications for knowledge representation and metadata management. In addition to standard recordkeeping requirements imposed on all corporations, as a

publisher the museum may have legal obligations to make its online presence accessible to the handicapped (in the United States under the Americans with Disabilities Act), or to protect minors from material that is considered inappropriate to them (in the United States under the Children's Online Privacy Protection Act). It is likely to have further obligations that are specific to its role as a collecting organization such as requirements to document the origins of human remains and grave findings (as in the United States under NAGPRA) (Grose, 1996), and will be required to document provenance of acquisitions of art that might have been looted or sold in contravention of laws protecting movable cultural properties (UNESCO convention and acts requiring return of Nazi- and Soviet-looted art). If it holds natural specimens, it will need to consider laws relating to endangered species and the documentation needs of the biodiversity community (see the Consortium for the Barcode of Life, http://barcoding.si.edu/index_detail.htm/). All these legal obligations imply requirements to represent knowledge about object provenance and interpretation in ways that support compliance with legal and reporting requirements.

Collections

In the 1980s, archives had substantial success in placing their metadata into library systems by describing their holdings "at the collection level" using MARC. Since then, EAD has been nominally accepted as a way of representing archival descriptions, which had a quasi-standard life as a prose genre, in XML. Museums also consist largely of collections, and some museums have been influenced therefore to represent their holdings in EAD. However, unlike archives, museum collections are largely artificial constructs (though some will have had an existence as collections prior to coming into the museum), and they do not have a pre-existing genre of prose descriptions to convert to XML. Representation of museum holdings at the collection-level is, therefore, not likely to be of any great benefit.

Public Interactions

Data placed on the Web will be interacted with, and can be made interactive or even open to community editing. Evaluating the interactions of the public with the data, and managing any annotations and uses made by the public, involves another layer of metadata. When the general public can search museum databases, museums discover the limitations of the data their collections systems hold, how inconsistently each facet of possible description is actually construed, and how much technical language is used in the data values. They often find that the public needs facets of description that are not usually employed by their curators. Some knowledge representations may be specifically oriented to the public, or even special age or interest groups within the public, while the same knowledge might be represented in a different way for internal use. For example, a botanical collection that curators use by searching scientific names of plants might need to be made accessible by common names, and an art museum collection which curators search by artist, title, date and genre might need to be made accessible to users who search for what they remember as the subject content of the images, something previously not catalogued at all (Bearman et al., 2005).

Because museum data on the Internet can enable two-way communications, museums find that they use it to build communities, which implies collecting metadata about possible clients. For example, the museum might create a Web form to gather memberships or reservations for a lecture series. Once the museum has experience with Web forms, they could inaugurate online forums, build collective documents, participate in interactive lecture dialogues or engage online chats. However, if museums seek to attract different audiences with their own specific requirements for museum content, and/or create mechanisms that permit people other than museum staff to add data to museum knowledge-bases, they will need to adopt sourcing for all their data. In other words, every piece of information will need to have metadata associated with it to say by whom it was created, when and under what authority (if any) and who owns it, and who can change it. No longer will it be acceptable that the contents of the museum databases speak "for the museum" and with that anonymous authority. Now it will be necessary for individuals to sign contributions to the database and speak with their own authority. By definition this reduces the abstract authority of the museum and brings it closer to the level of other institutions which can then articulate their views more equally.

Publication and Its Representation

As soon as the museum becomes actively involved in recruiting public attention, its publicity machinery must ensure that the public knows enough to participate in its programs. Information about where publicity has been released and what kinds of data have gone to which newsfeeds, must be kept along with the various releases. Relevant information, at a considerable level of granularity, must be keyed to where it has been made publicly available and to whom it is targeted. One potential strength of online information provision and online exhibitions is that the statistical preferences of prior visitors can be useful guides for subsequent clients. In order to construct systems that reflect what others have selected, whether using crude counts or sophisticated profile-based preference weighting, each "page" of display information, even those generated on the fly in response to queries and user-profiles, must contain its own history of use metadata. Recommender systems are constructed on such retained links and nodes. They can both support museum evaluation needs and assist users to follow paths that others have found useful.

Location-Based Knowledge

Once they are liberated from wires, users could be inside or outside the museum and their location, their paths, and even the direction of their gaze, become meaningful criteria in judging what information would prove relevant to them. Museums will need to make interpretive information sensitive to the location of users in order to meet wireless needs. Once knowledge representation begins to consider the location of the user, it adds value to reference the geo-location of a variety of facts in the life of the objects themselves—their creation, acquisition, and exhibition. This in turn makes it possible to deliver to users in those locations information about the events in the life of objects in a potentially remote museum. Knowledge representation that takes advantage of the spatial location of the user and the object throughout its life cycle will make museum information more relevant.

Experiences

Museums are dependent on making museum visiting a life-long habit, reinforced by good experiences. Museums are therefore at least as interested in building loyalty as any brand would be, and are beginning to realize that brand loyalty is reinforced by their using knowledge of their customers to satisfy known needs. Online visiting is no different, and in fact there are many advantages in the online environment since it is relatively easy to keep data about what people do and tell us, to recognize the customer instantly on arrival and to provide feedback about matters of interest that might have unfolded since their last visit or advise them about what other visitors with similar interests have been doing. Of course any such program requires that the museum create and maintain knowledge about its customers. This knowledge can't remain within the museum shop or files of the development office either; it must be fed into the delivery of online and onsite interactive and interpretive experiences. The knowledge representation requirements about customers can be considerable—involving tracking user behavior through exhibits and interaction events and building profiles that can individuate future experiences.

Multimedia Assets

As discussed elsewhere in this volume, metadata regarding surrogates has a growing place in museums. Data on the Web is multimedia, and the multimedia content is not just still images and graphics linked to text, but increasingly includes multi-modal, time-based data, such as animations, sound, video, and complex games, that have an elapsed performance time. The absence of universally accepted standards for encoding of multimedia has led museums to keep the same data as multiple MIME types. This, and the growth in absolute numbers of files, has resulted in implementation of media or asset management systems with huge overheads of metadata about media objects added to the knowledge representations. In the evolving jargon of museum systems, this is reflected in the use of the terms "content management systems" for the Web presence and "digital asset management system" for the in-house data objects in addition to "collections management systems" which keep track of the "real thing" as well as its surrogates and occurrences of its representations.

As the museum's investment in creating rich data grows, its need to keep that information available and accessible increases over time. The museum must consider life cycle information asset management as a priority, and this in turn means further investments in knowledge representation. If the museum obtains an image of a work in the collection by photographing it in response to a request, it needs only concern itself with keeping the request it received. But if the photograph is to be kept and digitized for future use, it will need to keep information about the digital asset (how it was made, how it is stored, etc.) in order to address version control and format obsolescence, along with the digital photograph itself. Archival issues, such as how long records are to be retained and under what authority they might be disposed, as well as metadata about format dependence needed to prevent obsolescence of formats while their content is still needed, must be addressed, and these in turn call for more knowledge representation about the authorities responsible for the content.

Sharing Knowledge

Museums have historically not been interested in the computer interchange of knowledge with other museums, though scholars and other cultural institutions see museum metadata as an attractive resource. Although many efforts have been made for twenty years, museums have not succeeded in finding a compelling reason to communicate with each other. Surprisingly, even the most obvious and best funded cases for benefits in such data interchange—the requirements of biodiversity researchers—has not generated implementations of museum to museum (M:M) data sharing. Specific projects to aggregate museum metadata have been forced to develop their own approaches (Bearman & Trant, 1998a).

While the objects museums hold are unique, and much of the knowledge that museums have about their collections and activities is likewise specific, an important component of museum knowledge is logically shared by other institutions. Indeed, the aspects of museum information that are shared by others are the most important properties from the perspective of anyone wishing to understand the world beyond the boundaries of the museum itself. It is through common data about people, places and things, methods and events that links between the particular and the more universal are made. There are times when it is preferable to link to, or incorporate, knowledge that is stored elsewhere rather than to replicate

it redundantly within museum databases. In these instances, structural mechanisms replace data content to bring in related facts. For example, if the museum acquired an object at an auction on July 17, 1953, the auction catalog might be incorporated by reference. If it was, the description of the object as it was known then, and as it was represented for sale, would add depth to the fact of the acquisition on that date. Rather than recording in museum databases popular explanations of the differences between lithography and other printing techniques, we could create a link to the Museum of Modern Art exhibition which is devoted to this topic (http://www.moma.org/exhibitions/2001/whatisaprint/flash.html). Of course, before the museum uses links for any information that it deems essential to its operations, it will need to ensure that the links remain active. Pointing, linking and incorporating by reference data that resides elsewhere is a method of knowledge representation that builds collective knowledge and links the particular to the more universal. The "info URI" scheme (http://info-uri.info/) developed to reference analogue object identifiers—such as ISBNs or LC call numbers—holds much promise for the many obsolete identifiers found in museum data repositories (National Information Standards Organization, n.d.).

Changing assumptions about the architectures that will support future information use have influenced the museum community, like others, in its choice of metadata packaging strategies. In 1993, the CIMI Standards framework proposed using a MARC-like interchange standard and querying remote museum databases using Z39.50 (Bearman & Perkins, 1993). This assumed that distributed resources would be brought together in central databases or that they would be queried in a targeted search. An alternative approach available at the time would have been to specify a standard structure for museum databases, which was rejected as impractical. Since 1995, arguments have been advanced for "industry-standard" cross database solutions like SQL and ODBC (Open DataBase Connectivity) as mechanisms for data interchange, though they don't address the need to agree to common schemas. CIMI (in collaboration with the Dublin Core Initiative) argued for extending the Dublin Core as a model both for content interchange and query, or "discovery," but they could not agree on which extensions would be required. These approaches should no longer be considered.

The preferred approach at present is federated architectures based on metadata harvesting using the OAI-PMH protocol (http://www.openarchives.org/OAI/openarchivesprotocol.html/), which the digital library community adopted in 2001. If the museum community was willing to adopt a basis of unqualified Dublin Core description, it could create implementation guidelines specific to its needs that would support federated harvesting using OAI-PMH and participation in harvesting efforts of other communities such as the Open Language Archives, or eventually those using LOM and METS standards. Tim Cole of the University of Illinois has been active in NSF and IMLS projects that are promoting this approach, but unfortunately little beyond conference presentations has been written that makes the rationale for this advocacy clear or demonstrates its utility.

The experience of the Web suggests that looser agreements between parties may still support a degree of integration that didn't seem practical without adherence to common standards in the past. But loose linking falls far short of interoperability. Metadata registries seem to be preferred by European initiatives, though the precise mechanisms by which they are supposed to work are still unclear (http://www.ukoln.ac.uk/metadata/). The W3C is promoting the virtues of RDF and the potential of the Semantic Web, which when deployed using a name space based schema declaration model does not require adherence to any given reference model. The Semantic Web model is appealing to some museums, though its full promise is still quite distant (Hyvönen et al., 2004).

Action-Centered Data Models

It seems likely that the Web will erode the splendid isolation of museums; if not because the museums find that they want to be part of a larger universe of information, then because the players in that larger universe increasingly appropriate museum knowledge and find ways to integrate it with their systems. Action-centered data models privilege a view that integrates resources of many cultural institutions (Bearman, Miller, Rust, Trant, & Weibel, 1999). By reinterpreting all facts as statements made about events, they unify time, place and ideas, the three remaining facets of Ranganathan's five elements other than objects and energy, and thereby link things with processes (Ranganathan, 1933). Together with the CIDOC Conceptual Reference Model (http://cidoc.ics.forth.gr/), this way of looking at relations as action linking entities emphasizes the unity in diverse objects that have been the subjects of actions such as discovery, invention, creation, publication, interpretation, analysis and presentation.

Although the development of an object-oriented version of the CIDOC relational data model was a logical next step for the ICOM Committee on Documentation to take, museum practitioners within that community did not take the lead in its development. Instead, academics took a lead, and ultimately brought the model to the International Standards Organization without much museum community input. Object-oriented models have had little influence on museum knowledge representation practices, but the CRM, and especially its support for event-centered views, has had an impact on knowledge models in the broader information community (Doerr, Hunter, & Lagoze, 2003; Hunter, 2002; Lagoze & Hunter, 2001). It may be ultimately that the model helps integrate views of heritage by serving as common frame of reference for understanding specific schemas (Lee, 2004).

CONCLUSIONS

Describing everything for every purpose and managing the data intelligently over time is not easy. It requires a great deal of self-consciousness about purposes. Explicit rationales for knowledge representation will be required to ensure others preserve the data properly. The ultimate payback will come when museum knowledge can be readily integrated with knowledge from other sources. This larger goal will require not just good knowledge representation practices, but the political and economic commitment of museums to cooperate.

Ultimately the environment itself will need to be built to support intelligent processes on the objects represented in it. We require a cyberinfrastructure of knowledge, with the toolsets and underlying methods to support connections between ideas and the objects that embody them. This vision is what drives the Semantic Web efforts, though I suspect that it may require less of a breakthrough in the means of representation and more investment in knowledge engineering.