Archives and Museum Informatics

Cultural Heritage Informatics Quarterly

In Memory of

Paul Evan Peters

Gentle Friend and Radical Visionary

VOLUME 10 • NUMBER 3 • 1996

Archives and Museum Informatics: Cultural Heritage Informatics Quarterly (ISSN 1042-1467) is published by Archives & Museum Informatics, 5501 Walnut Street, Suite 203, Pittsburgh, Pennsylvania 15232-2311; (412) 683-9775; fax 412-683-7366.

The journal is edited by David Bearman, whose authorship may be presumed for all items not otherwise attributed.

Archives and Museum Informatics carries news, opinions, and reports on information technologies, techniques, and theories relevant to archives and museums.

Submissions of press releases, publications, and software for review, articles, and letters to the editor are welcomed. Copy is preferred double-spaced. Longer articles will be requested in machine-readable form if accepted for publication. Deadlines for contributed articles and press releases are the 15th of March, June, September, and December.

Subscriptions are available on a calendar year basis at US$90.00 for institutions, US$50.00 for individuals (paid in advance, by personal check, and delivered to their home address), US$25.00 for individuals employed with subscribing institutions (mailed to business address), with a surcharge of US$10.00 for subscriptions outside the United States and Canada. Back issues are also available at a reduced price. All payments must be made in US currency. Credit cards are not accepted.

Archives & Museum Informatics also publishes occasional technical reports available for purchase. Prepaid orders include handling fees. Billed orders are subject to a US$5.00 billing/handling fee. Current titles can be obtained from the above address.

Archives and Museum Informatics
Cultural Heritage Informatics Quarterly

Volume 10 • Number 3 • 1996

CONTENTS

Asynchronous Mass Telecommunications

Until recently, asynchronous mass telecommunications was an oxymoron. Yet, in a very few years we'll look back at the twentieth century experience of broadcasting and think of it as an early stage in the evolution of mass communications, and our children will wonder what it was like to be limited by the fact that the audience was required to experience the communication at the time it was transmitted or monitor it to be played back at a later time. What is becoming clear about the evolving environment of Internet based (digital) communications is that it will soon have carry the modalities of radio, television, telephone and video-phone, have the combined penetration of personal and mass communications systems, and that every individual will be able to do something different from his neighbor at any time.

A year ago I would have challenged these statements on the grounds that the Internet was not and might not become ubiquitous, or the grounds that the experiences delivered over the network were still far from real-time video and interactivity, or on the basis of the paucity of proof that popular culture vehicles (including televsion) would be interested in delivering content over the network. This year, in the wake of the success of web based advertising, after the rise of Java and other interactive tools, and in light of the phenomenal growth in connectivity worldwide, there is little doubt in my mind that we will see the ascendancy of international asynchronous mass telecommunication within a few years.

This has tremendous consequences for content providers (read cultural repositories as well as newsletter publishers). In October, the Library of Congress redesigned its homepage following the

conclusions of the Library's Executive Committee that "the Library's Internet presence is now the major publication of the Library of Congress!"

My personal response to this new reality has been to move the reportage function of *Archives & Museum Informatics* out of the print medium (this is my last print issue with news-like reporting) and to organize a conference on Museums and the Web (March 16-19, 1997, in Los Angeles). The serious challenge will be to develop strategies for archives, museums, libraries, zoos, parks and other cultural institutions to pursue over the next few years. I am convinced the effect on cultural repositories will be the biggest change in the function and the popular perception of their function since they became part of the institutional landscape of the modern world. Although this is my parting editorial in *Archives and Museum Informatics*, I hope I'll continue to be around to comment on where we are headed and to think about how to get there. For starters, I think that in the new environment we will need to:

- seek a new economic basis for information delivery and public programming
- nurture partnerships with organizations which can advance our mission
- collaborate with like institutions to re-invent the social roles we play
- learn from those whose have grown up with this medium - they will shape its use
- experiment shamelessly knowing most experiments will fail

Have a good time. If it's not fun for you, you can't expect it to be much fun for visitors. The future of asynchronous mass telecommunications is ours to make.

Item Level Control and Electronic Recordkeeping[1]

David Bearman

Introduction: Archival Principles and Collective Control

Archivists are keenly aware of the huge volumes of records created by the day-to-day activity of a modern organization. They understand both that records are essential to the conduct of all the business process of the organization and that the vast majority of records should be disposed of when no longer required by the specific activity for which they were created. Ideally each record should be kept only for the amount of time deemed necessary for corporate purposes but be immediately available during that time as evidence of the transaction for which it was created. Practically speaking, this has meant that records managers and archivists have had to make sure that organizations segregated records according to the length of time they were expected to be of continuing value.

1 The author wishes to thank participants in the Monash University, 'Managing the Records Continuum' workshops, held in Melbourne and Canberra in June and July 1996 for their help in forcing the clarification of ideas in this paper, and in particular to Sue McKemmish, Frank Upward, Barbara Reed, Chris Hurley, David Roberts and Adrian Cunningham for reviewing earlier drafts and for their cogent criticism. My thanks also to Lisa Weber for a critical review of the penultimate draft which I believe helped to greatly improve it. A version of this paper was presented at the annual conference of the Society of American Archivists, in San Diego, August 29, 1996.

In the paper world, segregation of records by the transactions which generated them has resulted in establishing discrete filing (recordkeeping) systems for the records of different transactions, and retaining these, under administrative controls and in proximity to where they are used, as long as they are serving a current business purpose. Records are arranged within these recordkeeping systems in the order in which the business process they serve needs to retrieve them and whole systems, or discrete physical parts of such systems, are retired to records storage, accessioned to archives, or destroyed, when their "schedule" dictates. For as long as the records are kept, the "original order" physically imposed by the office filing procedure is the basic method for supporting access within the recordkeeping system, a valuable clue to the meaning of records and an essential component of their evidential value.

This approach to records is neither necessary nor desirable for electronic records.

The essential difference between electronic and paper records is that the former are only logical things while paper records are usually thought of only as physical things.[2] Physical things can be stored in only one place and in one observable order; logical things can be physically housed in many places but seen to be together. They can appear to have different arrangements depending on the views accorded to their users. In other words, the properties of logical things are associated with them through formal, defined, logical relations while the properties of physical things are associated with them as material objects with concrete locations, attachments and marking. Archival and records management methods have been developed to manage physical things, which

2 Sue McKemmish in "Are Records Ever Actual?", The Records Continuum: Ian Maclean and Australian Archives First Fifty Years, ed. Sue McKemmish and Michael Piggott, Ancora Press, Melbourne, 1994, pp. 187-203 has recognized that paper records are also, and should be understood as, virtual things, but this sophistication is rare indeed.

has both limited and, in some cases, simplified them. Archival theories have been developed to validate these practices and are therefore based on the assumptions inherent in managing physical things. As a consequence, archivists have elevated pragmatic responses to the nature of physical things to the level of ideology. The more we examine electronic records, the more instances of such an elevation of practice to theory are revealed.

For example, archivists have traditionally insisted that the internal arrangement of records within series must be retained undisturbed from the way the records creators kept them. The theoretical importance of this was that the order in which recordkeeping systems retained physical records dictated how they could be used within the office of origin and therefore provided evidence of the conduct of the business processes which created them. When records were re-filed into some order considered more "convenient" by a subsequent custodian (as when an alphabetical or chronological order was imposed where none previously existed) the evidence which record order had conveyed about the business process context and the methods of retrieval supported by the original office was lost. Even re-filing correctly, items that were misplaced in the original files, distorted the record and could have implications for inferences made about the records as evidence. What the physical order was preserving was the logical associations between records, but this was not understood.

In electronic environments the methods by which the originating office can use the records are not a reflection of the physical storage order but are instead established by the capabilities of the software environments in which the records are used. These software functionalities are likely to change over time. The capabilities of any given individual within these systems is further determined by the permissions and views accorded to those individuals in different relations to the records and these also change over time and with each user. Finally, the ways in which records are "filed" depends on the assignment (or lack of assignment) of

data values or on structural links defined in software architectures. Because the way that the records are organized on any storage device will not give evidence of their use or the business processes that employed them we must rely for such evidence on metadata (information about information systems and business processes) created contemporaneously with the record and its interaction over time with software functionality and user profiles. Thus the principle regarding original order is revealed to be that we must document the context of creation and use, including the logical associations of records in recordkeeping systems, in order to understand records as evidence, not that we need to literally "preserve" the original order of the physical records.

The concepts of series, record groups or fonds themselves are not truly physical but instead are logical associations. Because it is often the case that physical series or physical systems correspond to logical series and logical systems, archivists and records managers find it convenient to hold on to their physical notions. But in an electronic environment, where the boundaries of physical systems need not, and often do not, correspond to the boundaries of logical systems or organizations, the series and the fonds must be recognized as classification schemes intended to convey information about the transactions by which the records were created and by which they were maintained. Not only can the electronic record not be expected to have the physicality associated with either record series or fonds in paper records, it is unnecessary to impose it. All the actions we would have taken "collectively" based on physical proximity of records in traditional recordkeeping systems, we can take logically, without such physical aggregation, if appropriate logical relations are documented at the item level.

Beyond the pragmatic limits imposed by arrangement of paper records, physical aggregation has reflected the administrative boundaries of custody because physical control dictated who could see records and use them, which offices had access, and

when records were retained and destroyed. The fonds reflected the ultimate legal and administrative responsibility for records and their recordkeeping systems. The procedures of this administrative entity were crucial to estimate the trustworthiness of the records inherited at a later date. The archival principle of *respect des fonds* is, therefore, a pragmatic response to the fact that physical records lived for many years within the custody and under the control of an active administration whose practices determined the value and legitimacy of the records. Knowing that entity and its processes enabled administrators to hold the proper persons responsible for recordkeeping and historians to judge the value of the remaining record.

These indirect methods are not necessary in the electronic environment where much more direct responsibility, process controls, and *respect des fonds* can be imposed by control over the record metadata and the way systems can interact with records based on that metadata. These methods of control are discussed in detail in the final sections of this paper. Before we address those implementation issues, however, it is necessary to examine the dysfunction's introduced into recordkeeping by the assumptions of paper-based records practices, particularly in its focus on the "life-cycle" of records, and to introduce a new set of requirements for recordkeeping grounded in the social and legal concept of evidence.

I. The "Records Life-Cycle" Metaphor and its Consequences

The framework within which archives have operated in North America has been defined by stages in the custody of physical records and their collective management. Simply, and a bit simplistically but still reasonably accurately, we can say that during their active life, records in North America are in the custody and under the control, of the records creator. During their inactive life, aggregates of records (often entire series but frequently accessions

within them) are transferred into the custody and control of the records manager. And during their archival life, these record series and the larger collectivities they comprise, come into the custody and control of archivists.

In the United States especially, there has been a firm demarcation between records managers and archivists, between active records and inactive and archival records, and between records in agency custody and records in archival custody. The history of these distinctions is grounded in methods developed by the newly established National Archives in the 1930's-1950's as it struggled to deal with the rapid growth in the size of the Federal record and the simultaneous collapse of agency record keeping regimes. The concept of the life-cycle became firmly ingrained in archival thinking in the 1960's, and the separation of the records management and archival professions became increasingly pronounced until ultimately the National Archives gave up records management when it became an independent agency in the late 1980's. The consequences were predictable because even in the paper environment, only good recordkeeping from the moment of records creation can ensure good evidence and good archival control. In the electronic recordkeeping environment, separation between stages in the management of records from the moment of their creation onward, will not be viable.

Traditionally, archivists and records managers could not keep track of each individual record in a paper system, nor provide access to specific records based on their contents or on the specific transaction in which they took part unless the recordkeeping system of the office of origin used that characteristic as the basis for arrangement or a secondary index for retrieval. Hence archivists and records managers schedule, appraise, accession or destroy, describe and retrieve collectivities of records, generally at the series level. Because this practice does not best satisfy many users, the recordkeeping professions have developed theoretical defenses for it, but it is preferable to accept the obvious - we manage paper records collectively because it is too expensive to manage them individually. In the case of the electronic record, the reverse is true. It will be both more efficient and less expensive to control and describe records at the item level from the moment of their creation than it is to try to carry over into the electronic environment the methods of the paper world.

These two general assertions about the need to integrate records management and archiving, and the desirability of managing electronic records at the item level, can be understood better by looking in more detail at the traditional life-cycle stages and how they are impacted by electronic records.

1. Appraisal & Scheduling

Our practice has been to schedule aggregates of records, based on examination of those records. When those aggregates were essentially homogenous, that is when the record series in which the records were kept consist of records of a single type of business transaction, this worked reasonably well. When the record series were in fact defined as a matter of convenience to cover a large variety of types of business transactions, it worked less well. Often, as with a so-called "correspondence" series, in which the form or genre of the communication, rather than the specific type of business transaction defines the series, our practice failed altogether to dispose of records at the earliest practical time or ensure keeping only records required for longer periods or archival purposes.

In the paper world we have blamed these problems on records managers. If the records managers set up record keeping systems to distinguish between records of different transactions and keep them in separate series, they could be appraised more accurately. But the records manager could justly reply that the physical filing series which are established need to support the on-going functions of the office of origin. Simply establishing new physical series for each type of transaction would often unnecessarily

complicate that business process and frustrate its retrieval and use of records.

In a universe of logical records we can easily have the best of both worlds. We can develop information systems which support a variety of views of those records to support different users in the organization since "arrangement" is a matter of logical views rather than a single ordering of physical objects. At the same time, recordkeepers can dictate that each record must retain information about the specific transaction by which it was created, thereby satisfying our evidential requirements. Instead of appraising records, we can appraise the requirement for documentation of the activity or transaction that generated the record. Appraisal takes place in the aggregate still, now focused on the actions not on the documents, but control is at the item level where the effect of the appraisal is recorded in each transaction record.

Electronic records will not satisfy the requirements of evidence, unless they are irrevocably associated with the information about the context of their creation and receipt and information about their structure which is adequate for their subsequent reconstruction. This information (called metadata, or information about the data) is captured at the item level. Based on knowledge of the specific types of transactions which generated records, we can appraise their long-term value. When a transaction of that type takes place, its record can be made to carry scheduling metadata. In this way, scheduled retention can be associated with individual electronic records, based on the nature of the transaction and business requirements for its maintenance, from the moment of their creation.

2. Disposition

At some point, the archivist or records manager needs to act on the scheduling to effect the correct disposition (accessioning or destruction) of the records. In the paper records environment, collective level actions at this stage are often frustrated by item

level exceptions. For example, a record series may be scheduled for destruction but individual records within the series may have been the subject of law suits that prohibit their destruction. In paper, the records manager must decide whether to separate a physical record from the series, thereby violating its integrity, or to leave it in place and create a problem in managing disposition later. Occasionally, a record series may be scheduled for archival retention but individual records within the series may need to be returned to their originators as proprietary commercial secrets or destroyed for reasons of privacy protections. Such differences in the needs of the series as a whole and individual records within it currently cause significant administrative headaches and are not easily accommodated, but they are also not rare events.

If scheduling metadata is kept with each individual electronic record, any given record can be retained or destroyed when scheduled without acting on others. Computing systems make it simple to "collect" all the items which are supposed to be disposed in a particular way and act on them even though they are not physically stored together. If individual records that normally would be destroyed based on the rules associated with them at the time of their creation are covered by a court order, for example, their identification in response to that order would prevent their destruction at the time the rest of the series was destroyed. Similarly, the rules associated with disposal could ensure that records could be easily separated from a series, as when contractual obligations require items to be returned to a commercial owner. The nature of the metadata link to the record is such that this separation could occur without necessarily eradicating all traces of the fact that the records which were returned once existed as part of the series. While the contents of records would have been removed, metadata reflecting the existence of a transaction could (and I believe normally would) be retained by the archival system.

3. Description

The practical response to providing intellectual control over large volumes of records accessioned from paper recordkeeping systems was to employ top-down, collective description of records aggregates. Archivist typically described record after they were accessioned, which meant that certain things about the records were known only at the collective level - such as their provenance, the order of their arrangement within the recordkeeping system (and from that, implicitly, what could be inferred about their use), and their physical location. Because these attributes were known about the entire series, collection, or fonds, they were recorded in finding aids at that level of description. The individual items (record and filing units), inherited attributes associated with the larger collectivity, which in turn was presumed to have the properties associated with any larger aggregates of which it formed a part, such as the collection or fonds.

Item-level methods of managing paper based descriptions would have involved massive redundancy and substantially greater administrative effort since all information associated with the fonds, series and file would have to have been to written on each of thousands of separate cards and file folders on which details about specific items were recorded. These detailed cards or folders would then be filed according to their item level content descriptions and the knowledge of their context would be carried by the redundantly recorded data. To some extent these practices were incorporated into registry office functions in an age of paper recordkeeping which passed in the United States more than fifty years ago, although it can still be found in some parts of the world.[3]

3 I was delighted to see a standard Australian registry file folder, pre-printed to support the recording of file and item-level "metadata", at our workshop in Canberra this year. It both concretely illustrated the practical problems associated with managing paper records at the item level, and at the same time served to identify the specific metadata items thought necessary for evidence by a traditional registry office.

It goes almost without saying that automatic context and structure description within the metadata of electronic records at the item level would serve user needs better than collective description. Item level information is fundamentally more valuable because it can generate more valid collective level data in addition to serving the needs of item documentation. For example, archivists have described the date span of a series of records from the earliest to the latest, as in 1953-1987. In fact, of course, the record series may have individual items dating from December 11, 1953 to February 7, 1987, or only parts of the first and last years respectively. In addition, it may contain no records at all which are dated in the twenty months between March 5, 1976 and November 14, 1977. Automatic analysis of item level electronic data might reveal that eighty percent of the records were created one nine month period. Many such important facts about the collection are completely disguised by the collective date range but would be easily revealed by data recorded at an item level within an automated control system. A researcher could even see a pictogram of the dates of the records when asking about the series as a whole.

Such item level description, even of such a simple element of information as the date of specific records, has not been a regular part of archival practice because of the expense of acquiring such data in a paper environment, not because archivists did not realize that researchers would find such metadata valuable. Experience with researchers in some eighteenth and nineteenth century manuscript collections and registry systems which provided item or file level documentation has proved the value of such detailed description.

4. Retrieval and Access Control

But even where item level descriptions have been available, as in some heavily indexed early manuscript collections, paper based description systems allow researchers only to retrieve the records

of a particular correspondent (or, if they were filed by correspondent, the researcher will have a burdensome task assembling all the correspondence written to a given person in a single week). Again, little needs to be said about how these difficulties are obviated in an electronic environment, nor about the benefits this would give to the researcher who, as always, is pressed for time and needs to use it efficiently. What might be worth noting is how much more efficiently this researcher could work. For example, carrying all the materials needed out to a researcher desk could be achieved more efficiently (without the airfare and hotel bill, with the opportunity for simultaneous use by others).

Item level metadata also enables us to provide or limit access to materials which, for reasons of security, confidentiality or privacy, can only be viewed by some people, at some times and with some content masked. In paper there is no easy way to manage such records with the rest of the series or to administer access control without item level review. For records maintained in electronic form, appropriate metadata provided at the time of records creation can establish conditions governing access to the whole or parts of a record and pertaining to different users in different ways, and automatically ensure that records "show" themselves differently, and appropriately, to each class of different users. Confidence that records can be managed in this way can be critical to deciding that they can be retained at all, because if records with proprietary or personal information cannot be assured of safety, they are likely to be destroyed while still of value for other purposes due to risks associated with keeping them over time in less easily administered access regimes.

II. An Item Oriented Approach to Managing Electronic Records

1. Requirements

Over the past several years, archivists have been seeking an intellectual framework that can dictate an effective strategy for electronic recordkeeping. During the past three years, I believe such a framework has been defined by me and my colleagues at the University of Pittsburgh School of Library and Information Science.[4]

We began by asking what archives are for, and answered that they preserve evidence of human transactions for reasons of organizational accountability and personal identity. We conducted a focus group of experts in archives and records management to define the attributes of evidential records. Then we sought "warrant" in the literature of law, regulation and professional best practices that could help us define the necessary and sufficient attributes of evidence and by analysis of such statements of "literary warrant" derived a specification of the attributes of "recordness" or evidentiality.[5] The two sources were found to be in basic agreement, thereby supporting a consistent statement of the functional requirements for evidence in recordkeeping.

The specification of these functional requirements defines twenty properties which are identified in law, regulation and best practices throughout the society as the fundamental properties of records. The literature of specifications recognizes the danger of

4 NHPRC grant (#93-030) "Variables in the Satisfaction of Requirements for Electronic Records Management" see http://www.lis.pitt.edu/~nhprc for a specification for the full requirements of evidence in recordkeeping.

5 David Bearman, Electronic Evidence: Strategies for Managing Records in Contemporary Organizations (Pittsburgh, Archives & Museum Informatics, 1994)

natural language because it is often ambiguous, imprecise and subject to a high degree of interpretation. In order to ensure that systems would be able to rigorously enforce the assignment of these characteristics, we expressed the functional requirements in formal English as "production rules" or logical statements of simple observable attributes.[6] Because this forced us to avoid the ambiguity usually associated with requirements expressed as prose, the production rule formalism informed the process of articulating the prose requirements in addition to being used as a representation mechanism for those requirements. It also allowed the specifications to be logically refined, in a top-down decomposition, such that the most atomic component statements of the specifications were, in principle[7], observable states or properties.

The necessary and sufficient characteristics of data purporting to be records is a concrete set of metadata which, when present, satisfies the specification. By the requirements of evidence, if this metadata is inextricably linked to, and retained with, the data associated with each business transaction, it will guarantee that the data object will be usable over time, be accessible only under the terms and conditions established by its creator, and have properties required to be fully trustworthy as evidence and for purposes of executing business.

The functional requirements for evidence in recordkeeping[8] dictate the creation of records that are comprehensive, identifiable (bounded), complete (containing content, structure and context),

6 David Bearman and Ken Sochats, "Formalizing Functional Requirements for Recordkeeping" unpublished draft paper included in University of Pittsburgh Recordkeeping Functional Requirements Project: Reports and Working Papers (LIS055/LS94001) September 1994

7 I say in principle here because this is not a finished and fully validated standard and has not been subjected to the kind of testing that would give complete confidence in the specific low-level observables. Since the discussion here is more about strategy, however, it should be stressed that as a strategy the method seems fully proved.

and authorized. These four properties are defined by the requirements in sufficient detail to permit us to specify what metadata items would need to describe them in order to audit these properties. This descriptive metadata cannot be separated from them or changed after the record has been created. Several additional requirements define how the data must be maintained and ultimately how it and other metadata can be used when the record is accessed in the future. The metadata created with the record must allow the record to be preserved over time and ensure that it will continue to be usable long after the individuals, computer systems and even information standards under which it was created have ceased to be. The metadata required to ensure that functional requirements are satisfied must be captured by the overall system through which business is conducted, which includes personnel, policy, hardware and software.

This specification for evidence serves to identify the data required for such purposes as were proposed in last year's draft NIST standard for a "Record Description Record" or the recent report of the Research Library Group/Commission on Preservation and Access Task Force on Archiving of Digital Information.[9] The approach has been embodied in independent proposals electronic recordkeeping made by Astra to the Swedish National Archives[10] and in proposals influenced by our work from an influential group of Australian archives[11]. In addition, we have seen direct implementations of our work in projects in Indiana, Vermont, and Philadelphia, funded by the NHPRC as follow-ons to our study.[12]

8 op.cit, fn.4

9 www-rlg.stanford.edu/ArchTF

10 Ulf Andersson, "SESAM. Philosophy and Rules concerning Electronic Archives and Authenticity" (ASTRA AB, 28 Feb.1996) 86p.

11 Australian Council of Archives, "Corporate Memory in the Electronic Age: Statement of a Common Position on Electronic Recordkeeping", May 1996.

It is important to understand that the requirements for evidence in recordkeeping are not the same as, the requirements for an electronic records management system. An application system will be implemented in a concrete place and time, and operated by real people in the course of specific assigned duties. As a consequence, requirements for system security (as contrasted with records security, or integrity), systems compatibility, interfaces and standards (as contrasted with records inter-operability and migratability), and support for concrete business processes of the records management function would need to be addressed in a comprehensive statement of the requirements for acquiring a records management system along with requirements of evidence in recordkeeping.[13]

We begin with a single, simple conceptual framework encompassing what constitutes a business transaction, evidence and an acceptable record:

12 "Functional Requirements for Evidence in Recordkeeping: Invitational Meeting. University of Pittsburgh, February 1-2 1996,"Archives and Museum Informatics, vol,.9#4, p.433-437

13 As an exercise in gaining a better understanding of the difference between requirements of the property of "recordness" and requirements of the application of records management, many of these further requirements were identified in the Monash University workshops in Melbourne and Canberra in June and July 1996. I believe this confusion between records management systems requirements - that is the requirements of an application system designed to support the functions assigned to records management offices and archives - and the requirements of recordkeeping, has confused the debate over many years. It was one of the difficulties faced by the SAA CART committee which was charged with both electronic records and applications technology, and it currently confuses the work of the Records Management Task Force of the U.S. Department of Defense, which is designing a records management application system but slips into trying to impose those requirements as the requirements for recordkeeping.

2. Transactions

Transactions (trans-actions) by definition are actions communicated from one person to another, from a person to a store of information (such as a filing cabinet or computer database) and thereby available to another person at a later time, or communications from a store of information to a person or another computer.[14] Because such trans-actions must leave the mind, computer memory, or software process in which they are created (or must be used, "over-the-shoulder" as it were, by a person with access to the same computer memory), they must be conveyed across a software layer, and typically across a number of hardware devices.

3. Evidence

Not all data that has been communicated or created by information systems in contemporary organizations is captured as evidence. Information systems are generally designed to hold timely, non-redundant and manipulable information, while recordkeeping systems (information systems designed to capture and maintain evidence) store time bound, inviolable and redundant records. Therefore, application environments that support the ongoing work of the organization frequently, or even usually, do not satisfy the requirements for creating evidence. Recordkeepers need to provide in-house information managers with a rigorous definition of the distinct requirements for recordkeeping. Without explicit and testable specifications, computing application and electronic communications systems will continue to fail to satisfy the requirements for recordkeeping and will be a growing liability

14 David Bearman, "Electronic Records Management Guidelines: A Manual for Development and Implementation" in United Nations, Administrative Coordinating Committee for Information Systems, Management of Electronic Records: Issues and Guidelines (New York, UN, 1990) reprinted in Electronic Evidence, op.cit.

to companies even while they are contributing directly to day-to-day corporate effectiveness.

4. Records

Records are at one and the same time the carriers, the products and the evidence, of business transactions. Any organization that wants to use electronic documentation as evidence needs to create records. Records oriented professionals within organizations, such as senior management, legal counsel, auditors, Freedom of Information and Privacy officials require records, not just information, to support their work. Business transactions must create records which logically are metadata encapsulated objects, although in our implementation model records need not be physically stored in this manner. In these records, the contents of the transaction would be preceded by information identifying the record, the terms for access, the way to open and read it, and the business meaning of the communication. Metadata encapsulated objects may contain other metadata encapsulated objects, because records frequently consist of other records brought together under a new "cover", as when correspondence, reports and results of database projections are forwarded to a management committee for decision.

Our concept of evidence makes it important to know when records were used and how, in what ways they were filed, classified and restricted in the past, and, if they have been destroyed under proper disposition authority, when and by whom that act took place. It is also important to know what redacted versions of records were released over time. Transactional data reflecting the history of its use (events in its life subsequent to creation), provides the documentation traditionally associated with archival description, but instead of such data residing only at aggregate levels, it is possible to define electronic records metadata structures that enable us to search for specific records based on information about the instance or concrete business transaction which generated them.

In addition to ensuring that the data we capture is a record, and can serve as evidence, metadata should be defined so that it makes data objects communicated across software and hardware layers (and therefore any communications over a network):

- self-documenting
- self-authenticating
- self-redacting
- self-migrating
- self-disposing

These properties, while important for simplifying the management of records (especially in an inter-networked environment in which hundreds of millions of records may be created daily), can be made direct consequences of keeping records if attention is paid to the structuring of the metadata that makes records evidence. Appropriate metadata can ensure a degree of software independence.[15] In addition, an ideal model would ensure that all record objects we create would be interoperable between record-keeping systems environments to give them independence of specific custodial settings. Furthermore, a system for metadata management which has appropriate modularity and content stand-

15 The actual degree of software independence that can be achieved depends on how long any given "standard" can be expected to remain a standard. In archival terms, this is often not very long. When the independence provided by standards expires, the fact that the data was recorded in a standard will usually provide a route to low cost migration, often directly into a successor standard. Many data objects we create today will not be standard and the metadata with which we label them must flag the dependencies of the data (including their dependency on standards) so that a future review of record headers can locate sources of brittleness and segregate records requiring migration to new software formats before they become unreadable.

ardization can support formally auditing the business system which generated the information object. It can enable the auditor to locate the transactions and the software, hardware, procedures and policies surrounding a transaction, to determine where they contribute, or fail to contribute, to the creation, maintenance and use of evidence. While no system of management can be self-auditing, a communications system built to ensure that appropriate metadata is captured for evidence can support a level of management accountability that it was not possible to implement or enforce in paper-based environments.

B. A Reference Model for Business Acceptable Communications

A rigorous technical standard is required if we are going to implement the functional requirements for evidence in recordkeeping within all electronic communications environments in the future. The goal of such a standard would be to make communications received over networks trustworthy for the purposes of conducting business. It would be designed to ensure accountability and protect organizations against the risks of loss of proof of their past behavior. As a consequence, it would greatly simplify:

- the management of huge volumes of communications from heterogeneous hosts,
- the proper retention and disposition of records,
- auditing the use of records for business, and
- the appropriate management of private, secure, proprietary or confidential data.

A side effect of such a recordkeeping standard is that it would enhance the business value of the data that it preserves. These business benefits would include:

- providing data for market and other research.
- documenting decisions, policies, events, etc.
- documenting R&D and other business related processes.

For archivists, the most important consequences of adoption this kind of standard would be that electronic communications carried on in the regular course of business would always be captured in a way that was:

- evidential
- cost effective to store, whether in centralized or decentralized repositories
- securely controlled against alteration from the moment of creation
- highly cost effective to migrate across software and hardware dependencies
- supportive of automatic management of future access and use, including disposition and selective release

In December 1994, I proposed such a standard in a draft "Reference Model for Business Acceptable Communications" (abbreviated as BAC).[16] The Reference Model recognized that while the metadata requirements for evidentiality or "recordness" are necessary components of business acceptable communication, and must be accommodated by any reference model, they are not the only source of functionality of such electronic records. Requirements of object standardization efforts designed to provide support for a system of access and use rights management, networked information discovery and retrieval and registration of intellectual property have led to elaboration beyond the properties identified as necessary for assurance of evidence. Specifically, over the past eighteen months I have modified the model to reflect requirements being addressed by other efforts to develop widely

16 David Bearman, "Functional Requirements for Recordkeeping: Metadata Specification" (Unpublished draft, 12/21/94)

applicable models for network metadata management, such as those which support:

- registration of persistent addresses for information resources and intellectual property[17]
- networked information discovery and retrieval[18]
- a system of access and use rights management, as required by electronic commerce[19]

As a consequence the model now supports an environment of "registration" services for unique domain identification, "resolver" services for dealing with terms and conditions of access or use, and information discovery and retrieval services. These extra requirements are seen as positive and supportive of the overall tactics of the BAC model because they validate the assumption and represent commitments by other actors to participate in the scheme.

The emergence of a class of electronic commerce applications based on encapsulated and encrypted objects and token interchange between resolvers and object users (such as the IBM InfoMarket launched in May 1996) further validated the assumptions I made in 1994/95. Not only are these classes of applications promising in themselves, I believe they are quite likely to be implemented in widespread network-based toolsets by Microsoft, Netscape, and others as well as to find their way into API tools and communications structures.

The proposed Reference Model for Business Acceptable Communications attempts to specifically address these additional requirements as part of a dialog that must take place between advocates of mechanisms to support these different fundamental purposes through an overall structure for metadata encapsulated objects.[20] It does so by clustering these metadata data categories and elements so as to achieve functional modularity[21] and then arranging them into six layers to support the technical processing and interchange requirements of a widely distributed networked environment. The six layers and their clusters are:

17 The Internet Engineering Task Force work on Persistent URL's, and work on handles as part of the Library of Congress Electronic Copyright Registration project as reported by Bill Arms of CNRI at CNI Spring 1995

18 For example, the Networked Information Discovery and Retrieval study directed by Clifford Lynch for the Coalition for Networked Information. See also the reports of the Library of Congress Electronic cataloging meeting in October 1994 and the results of the Dublin, Warwick and September 1996 Dublin-image metadata workshops.

19 Such as the IBM InfoMarket_ Cryptolopes_ , which have recently announced plans to use Xerox's "Digital Property Rights Language (DPRL). Other commercialization's have been announced, but not yet launched, the EPR. For discussion of these and other encapsulated object based intellectual property mechanisms see their web sites, or those of ELSI. In addition, see articles by John Erickson (Cornell), also easily accessed on the www.

20 see, David Bearman and Ken Sochats, "Metadata Requirements for Evidence" at www.lis.pitt.edu/~nhprc/

21 This concept refers to the fact that each group of metadata elements performs a specified task, and that these tasks are logically required to be performed in the order in which the metadata clusters appear, hence the identification of an object, the establishment of its relevance, the determination that the user has rights to access and use it, the decoding of its structure, and reporting on its context, all take place prior to the presentation of its contents to the user.In certain areas, particularly regarding structural dependencies of data objects representing non-textual content, we have specified a potentially extensible set of modality specific data elements by naming a metadata category but not identifying specific metadata fields/elements within that category. This reflects the recognition that we can never completely specify the data that will be required to document the structural dependencies of future data types.

Layers and Data Clusters in a Proposed Reference Model for Business Acceptable Communications

- Handle Layer
  - Registration Metadata/Properties
  - Record Identifier
  - Information Discovery and Retrieval
- Terms & Conditions Layer
  - Rights Status Metadata
  - Access Metadata
  - Use Metadata
  - Retention Metadata
- Structural Layer
  - File Identification
  - File Encoding Metadata
  - File Rendering Metadata
  - Record Rendering Metadata
  - Content Structure Metadata
  - Source Metadata
- Contextual Layer
  - Transaction Context
  - Responsibility
  - Business Function
- Content Layer
  - Content-Description
- Use History Layer

III. Implementing Item Level Control in Electronic Recordkeeping

However attractive item level control might be in principle, archivists and records managers need to understand how it will work in considerable detail before they are likely to commit themselves to trying to bring it about. The following discussions of implementation approaches, and options, therefore addresses how:

1. business transactions, retention requirements, and structural requirements are documented,
2. metadata and records are stored,
3. records are uniquely identified,
4. disposal and migration are managed,
5. a record is captured from the business information system to the recordkeeping system,
6. access control and redaction are imposed,
7. the records are retrieved and delivered to users, and
8. a history of use is kept for uses which require it according to business rules.

1. Capture

A trans-action is communicated from one physical or logical place to another, whether it is from one person to another, one hardware/software machine to another, or both. As such it crosses a logical switch, and when it does so, it can be captured. What a business considers a transaction, we have called a "business

transaction" and the Swedes have more recently dubbed a "causa."[22]

Electronic recordkeeping require us to distinguish between computer transactions and "business transactions". Most existing information systems are designed to update computer record for software transactions which have no business meaning, such as background saving of a file on which someone is working or during the spelling check through a long document, but will typically not create a record of common business transactions which do not change data in the system. Yet some such transactions, such as querying a decision support database, probably do require evidence under most definitions of what constitutes a business transaction. Implementations will need to impose the concept of business transactions, rather than that of systems transactions, on their environment.

Every time a business transaction crosses such a 'switch' implementers will want to create a record of the transaction. This record will consist of the content of the transaction encapsulated with metadata, while allowing the data and systems instructions created by the application to be communicated within the information system where it will do the work of the application and be available for further manipulation. In other words, the data in the information system continues to act in the way the application designer intended (updating databases, being available for users to store as information copies, etc.), but from the perspective of the recordkeepers, all data resident in application system becomes a convenience copy, rather than a record, and can be modified under the rules of those systems because the record exists elsewhere, as a separate object, which is not subject to modification.

When users generate a "Business Acceptable Communication", consisting of content encapsulated by all the metadata

22 Ulf Andersson, op.cit., fn.11

necessary to ensure its integrity and longevity, the record should be split off from the application systems environment and sent to a separate recordkeeping system or API layer recordkeeping service where it will be kept intact. This means that systems implementers need to construct 'traps' in which they can capture the business trans-action along with the metadata required for evidence. Most of this data, such as the time of the transaction, the identity of the sender and recipient, and the structural dependencies of the data, can be readily adduced from information available to the application and operating environment. The issue is how to generate, and capture, the metadata which identifies the business transaction-type or task of which the record is evidence.

In structured applications, each application system task can be identified and appropriate metadata defined for any transaction resulting from the task. Capture can easily take place using escape code sequences attached to each application system task. In unstructured business uses of application "utilities" such as word processors, email systems, scheduling facilities or spreadsheets, identification of the business transaction in which users are engaged is more difficult. Implementers will require some cooperation of users, although they can enforce this cooperation and make it minimally intrusive if they are clever. The basic strategy is to capture the requisite metadata items by assigning them to forms, style sheets, distribution lists and other objects created by the application software and used for specific business purposes. Users are then encouraged, or required, to employ the relevant "style sheets" and other settings in the conduct of what was previously unstructured activity.

At its most permissive, implementers provide users with value-added functionality launched by business process icons located in the user interface along with the familiar software application icons. The choice of business-process methods would launch capture routines for the resulting transaction and users would be enticed to use the value-added services by their benefits. Fortu-

nately, 1996 seems to be the year in which workflow management tools with object oriented metadata assignment finally come into their own which, for archivists, means a variety of off-the-shelf applications that can be set into user interfaces between users and the software environment in which they work. These interface managers can make the off-the-shelf application software appear to be a series of business applications and to label the records communicated by conducting those business transactions according to the retention, indexing and access requirements of the underlying business requirement. Prototypes of such implementations, which can even be placed over legacy systems, have been done for John McDonald at the National Archives of Canada, and by many others.[23]

Sophisticated approaches to "automatic" metadata capture would provide icons representing the business tasks in which a user may engage, based on process data models and business rules of the organization, rather than icons representing software applications; user selection of tasks would assign metadata to the objects created by the application. For example, a manager drafting a "directive" would open an icon for "Directives" rather than for "word processing". The "Directives" icon would run a configuration "client" designed to open appropriate software applications. The client executes a kind of "macro" which configure the application software in a way that utilizes its style sheets, self-documenting features, views, device drivers, etc. for the particular business function in which the user is engaged. Thus, in the case of our "Directives" client, it would call up the "Directives" style sheet. When the "Directive" is sent, it would pull up the correct distribution list from the company databases and send the directive by email, fax, internal mail, etc. based on settings in that database.

23 See, John McDonald, SAA 1996, report on the prototype developed for the National Archives of Canada by The Workflow Automation Company Inc., Toronto which follows essentially the same principles I used in the designs for the RLG AMIS Project in the early 1990's.

It would automatically schedule retention, file the directive in the organizational directives series (perhaps indicating the obsolescence of the prior version) and otherwise execute the business process needs and rules that should be imposed based on the "Directives issuing" process.

Such "clients" also provide the metadata necessary to identify the business transaction when a record of it is created. In a more rigid implementation, we might allow access to application software functionality only through software clients launched by icons in the user interface. In either case the transactional locus metadata, and metadata dependent on that information, such as retention period, access and use restrictions, filing rules and structural metadata are embedded in the selection of the business transaction icon/settings without explicit definition by the user.[24]

24 Other, more complicated or less precise, methods of identifying transactions which are the source of records have been proposed. Some thinkers have argued for an artificial intelligence approach in which an 'evidence service' in the Application Platform Interface which would capture transactions based on a knowledge-base of organizational communications content, form and addressing. Some variant on this would be required if all transactions are not considered records, which is a major reason I've argued against this concept as articulated by NARA. In principle, such an intelligent service could analyze transactions and assign them metadata attributes required to ensure their authenticity and survival but in practice, a service would need significant knowledge of the rules of communication within a particular business so as to identify transactions of specific types and adhere to the appropriate retention periods, access and use rights, and filing rules and studies to date show little likelihood that this level of knowledge is available or could be implemented in a rule base. At one time I thought information systems staff could identify components in the systems architecture, from storage devices serving as corporate file rooms to telecommunication switches linking to other LAN's, WAN's or systems, and have metadata attributes assigned to records based on where they originated, to whom they were communicated, and the technical characteristics of the transmission. While this might be possible, it also requires a significant analysis of transactional traffic and is susceptible to collapse as the characteristics of the content change (which they will). Finally, in conjunction with corporate policy and procedure individuals could be required to complete document

A variant of this option is being employed in "Intranet" implementations in which corporate users employ functions provided by the action office to make requests for services. Because the action office defines the types of requests, it can embed metadata into the resulting records. Likewise, in a distributed filing environment, records filed in certain places and under particular headings would be given metadata attributes upon arrival at the filing server application. Records deemed to be lacking appropriate metadata to leave an organizations' boundaries, or even to pass outside the LAN serving one work group, could be assigned those attributes or be returned to sender to provide the necessary descriptors.

2. Documentation

Implementing electronic recordkeeping means ensuring that metadata associated with records provides adequate evidence of electronic transactions at the level of the specific transaction within a defined business process. As such, recordkeeping systems need to interface with business process models to capture business transaction identification data. Some of this data, as discussed earlier, can be obtained from users when they sign on to the system, while other information such as the identity of the business task, authorizations, and the terms and conditions associated with the record of the transaction, must be brought into the system by actions taken after signing on, as discussed briefly above.

It would be possible to design application software that recognizes, or could be set to accept, definitions of business transactions boundaries but the differences between organizations would likely make implementing such software complex and maintaining its knowledge of local business processes costly. Already, organizations have certain parameterized features of application systems that can be employed to ensure the satisfaction of some of the functional requirements for recordkeeping. For example, word processing systems can support corporate record creating requirements if the users of such systems exclusively employ style sheets defined in such a way as to distinguish between transactions based on their process location and business purposes. Geographic information systems often have reporting features that allow the user to create output files of all the relevant layers of data incorporated into a query response. But these are rarely implemented due to the human costs. Instead what is needed are automatic means of labeling records with contextual and structural metadata.

However we go about it, we will want to document four aspects of the record at the time of capture:

First, we want to assign as much as possible of the contextual metadata to a transaction record based on knowledge given to the system by the user during the routine process of getting to the position that the transaction can be executed. This means capturing who the user is (and hence the full organizational context) through user sign-in. It means capturing what the user is doing (and hence the full procedural or business context) through a combination of the system functions chosen and the users explicit selection of a business process in order to facilitate its execution.

Second, we want to assign as much as possible of the structural metadata to a record based on the choices we have made about the format in which to capture records. The relevant structural definition is not the definition of the dependencies of the application environment itself but rather the dependencies of the recordkeeping structure. Thus we could save records from a proprietary application systems in a widely accepted standard, such as SGML, and capture all relevant metadata simply by recording that, and any publicly registered DTD that was being followed.[25] If the

profiles as part of routing and filing transactions, but such intrusive approaches tend to be resented, and often subverted, by users and should be avoided.

25 The reference here to DTD's (Document Type Definitions), which could be

applications software we are using does not offer the option of saving in SGML, at least we could make a record in RTF or PDF (or the most robust standards to which our current application could natively write). In all cases we would document the dependency on the standard, in the record metadata.

Third, we want to capture as much of the content metadata as possible through automatic means and make sure it is a independent of applications as possible. By using features of the application environment, for instance "declare" type functions by which one application informs another of its content structuring rules (data definitions and variable values), we can get the system to incorporate content metadata in the content layer.

Finally, we need to generate the business rules for keeping, providing access to, and managing these records over time and make sure that they are executable. Some of this information, such as the retention rules which govern keeping and destroying the record, can be made to come along with the identification of the correct transaction if we build business process models with appropriate retention rules linked to their records. Other information, such as the presence of proprietary, private or secret information, cannot be linked simply by the kind of transaction (although some transactions cannot have some of these classes of data in them). Here we need rules by which to recognize, and flag,

any kind of registered data set whether EDI, a MARC record format, or otherwise, reflect my belief that when the metadata needed by a specialized domain has an essentially application related purpose, but is not required for recordness, it is preferable to satisfy this application purpose by definition of an interchange format or inter-operability model. The interchange standard can be cited in the metadata for Business Acceptable Communications and the data content can then be opened by knowledge of the requirements and structures of the standard without further elaboration. This has the dual advantage of efficiency of definition and ease of migratability as all records corresponding to a specified protocol can be re-presented in a new standard if the old format is superseded.

portions of records that require special treatment, or authors would need to do this as part of the initial business transaction.

These steps in capturing metadata, together with metadata inherent in the communication process such as the date/time and identity of the recipients, makes up all that is required to ensure the evidential character of the records at the time of their creation. It should be noted that not all the metadata required for recordness needs to be recorded in detail in each record. For example, audit findings pertaining to compliant organizations and accountable systems are themselves business transactions of the organization. Transactions created between such audits need only cite the audit (e.g. meta-meta-data) to document that the system which created them was compliant; they need not carry all the metadata of such an audit. Similarly, much detailed structural information is contained in volumes of technical compliance tests for standards. Records conforming to standard structures may therefore be documented by reference to such external standards and need not carry all the metadata items required to specify the standard itself.

3. Storage

To be evidence, records must be inextricably linked with their metadata and inviolable in their content for as long as they are kept. Where the are, physically, is irrelevant as long as they are properly protected and controlled.

No specific computing model must be employed in the maintenance of recordkeeping systems although it may seem that the discussion of communicated transactions to this point has used the terminology of object orientation. The content of the record need not be physically stored in the same place, or same computer record, as the metadata, but storing all of the metadata with the record content in one encapsulated object so that metadata is always stored and transported with the record simplifies the long-term management. The encapsulation approach also has the advantage that a record, when retrieved, is physically self

explanatory. A perceived disadvantage of encapsulation is that a considerable amount of redundant metadata is stored with each record, adding to the overhead associated with every action taken. However, models of likely metadata content which I have developed strongly suggest that this overhead will, in normal business environments, be trivial with respect to the size of the data content. I believe the advantages of not having to worry about integrity of pointers in separate structures will make encapsulation a better option than keeping the redundant data in separate, relationally linked, files.

Nevertheless, records can also be stored in standard relational, hierarchical or even flat database management systems. While this approach avoids the overhead associated with communicating records, it requires more sophisticated management over time since assurance must be provided that neither the record, nor its associated metadata, could have changed. If we store the metadata elsewhere than in an encapsulated object, it will need to contain a pointer to the content and a signature (hash) of the contents, and the content will need to contain a pointer to and hash of the metadata. Even with such cross pointing and cross hashing, database integrity and security will need to be ensured by the system over time as part of the control environment. Some implementations to date seem to be choosing a tactic between these two extremes.[26]

The model itself allows for either centralized or decentralized custody. The declining cost of decentralized storage and the rising cost of central backup and hot site management suggest however that the solution of centralizing custody in an electronic environment lacks the major benefit associates with centralization of storage in paper-based environments. My findings in contemporary organizations are that unused storage on desktops far exceeds

26 This is explicit in Philadelphia, see Archives and Museum Informatics, v.9#4 p.435

the capacity (already nearly full) of central storage services. It is possible to develop corporate recordkeeping strategies that use such completely distributed custody by taking control over unused desktop disk space for corporate storage. In such a model, records must be protected from destruction by ensuring that they always exist in multiple copies, preferably in topographically independent LAN's, so that the disappearance of one of these copies from the network (as when a local machine has been brought down for servicing) automatically results in the creation of another copy somewhere else. In most topographies, three copies kept in systems that are independent of each other and are not susceptible to the same natural disasters or human interventions, with an object status monitoring directory in a fourth location, would ensure protection equivalent to that of central storage with off-site backups and a hot site, at a fraction of the cost.

Of course, records also must be protected from change. Regardless of how they are stored, only copies of records should be given out to other systems, and as soon as they are opened they need to lose the validation bits which certify their recordness. These validation bits could be the leader of the encapsulated object which is managed by its "read" applet and/or a "certification" of authenticity given out by an external agent. The same function can be performed by cross-hashing records stored in linked relational tables; each part of the logical record which is stored in a discrete physical record, needs to carry the hash of the content, and the content needs to carry the hash of the metadata. In this way neither can be altered without the fact of alteration being apparent.

4. Identification

One of the potential advantages of item level control and especially of metadata encapsulated objects all of which conform to the Reference Model for Business Acceptable Communications, is that records from everywhere, including records from more than one records creating organization, can be stored to-

gether in a simple and uniform recordkeeping environment. Therefore, it is a major concern how the record identifier uniquely assigned by one domain is guaranteed to be unique when the object is incorporated into a universe in which identifiers assigned by other domains are present. We know that uniqueness can be ensured by combining a unique identifier within a domain with a unique identifier for the domain. Practically speaking, however, how can we ensure that domain identifiers will be truly unique to a person or organization? This issue is being addressed by the Internet Engineering Task Force and others who are assessing schemes to "register" domain identifiers, or issue them without serious overhead. Because billions of unique business transactions will flow through worldwide communications systems within and between organizations and between individuals and/or computers daily, it must be possible to uniquely identify them all.

Below the level of domains, there are two different strategies for identifying transactions. We can give transactions an identifier based on an analysis of the business context in which they were created, or we can assigned them an arbitrary identifier, consisting of a sequence number. Arguments can be made in favor of either approach, although I believe that analysis of business function, process, activity and task, which is necessary in any case to appropriately document the transaction, should be encoded here so that retrieval of objects by functional provenance, which will be quite important during the active life of the record, is facilitated. In either case, a time stamp will ensure complete uniqueness and assist in subsequent retrieval.

5. Management

Because we have found that the most practical approach to management electronic records is to capture contextual metadata at the item level, terms and conditions for access and use can be determined either for all instantiations of a particular type of transaction (task) or for a single item, and will be enforced at the

item level in either case. For example, retention rules will be the same for all transactions of a particular type, but disposition can be automatically determined based on the specific date or recipient. Because every transaction creates a record, we can obviate the need for such costly processes as item by item review of case files in order to remove records used in a subsequent transaction, such as a court case. Instead we destroy all the original records as intended by their retention schedule, secure in the knowledge that a record copy is incorporated within the transaction that reviewed the files under the discovery process or in the management meeting which considered them. If the use of appropriate style-sheets or forms is enforced in business transactions, we could not only further segregate out of a series those records involving attorney client privilege, or containing confidential medical or proprietary information, but also provide for automatic redaction of records based on the profile of the potential future user.

Archivists and records managers with whom this model has been discussed extensively during a recent series of workshops in Australia tended to agree on a set of business rules for recordkeeping that would be likely to be put into effect as methods governing metadata encapsulated objects (MEO's). These included:

- when records are opened in the course of a business transaction, the method makes a convenience copy lacking "recordness" bits and retains a full "evidential" copy with "recordness" bits for incorporation into the resulting transaction (this was called big fish contains little fish).
- when records are "deleted" under records retention schedules, the contents of the record and the structural metadata and terms and conditions of access and use are destroyed, but the handle, context and use history are not, and a final transaction is added to use history to document the rules under which the disposal took place.
- when records are incorporated into other records, the terms and conditions for disposal of the parent record govern the

incorporated records, but the terms and conditions for access and use in the original records still apply to their use within the subsequent transaction.

- when records are released under restricted terms and conditions, either because access to them is limited to a specific class of people or because view or use restraints are placed on released copies, the use and user are recorded in the use history of the record being released in addition to the actual content released being incorporated into a new transaction record.

As a practical matter, we must also develop means to monitor metadata values in order to make the necessary software migrations at appropriate times in the life of records. If the records are kept as encapsulated objects, a secondary index may be desirable. Of course the most important management issue faced in migration is not just to migrate records to new structures before the old ones are no longer supported, we need to make good decisions about logical mappings in order not to introduce too much noise with every migration and ultimately lose the message in digital copying as surely as with did with multi-generational copying of analog messages. Needless to say, some people also worry that these software migrations, if they continue to need to be done as often as once a decade or more, will become too costly to support and that as a consequence some records of value will be abandoned. Within the environment in which recordkeeping takes place, stringent approaches to configuration management will be essential to ensure that record documentation retains critical usable metadata.

It is noteworthy that the proposed approach to archiving and maintaining business acceptable communications does not require us to include information about physical formats and media within the record metadata. Rather the environment in which records are kept will need to be one in which managers move data from one medium to another as required to assure backup and preservation

of the data. Because it is presumed that media that are currently supported will always be used and that data transfer to current media will take place in the normal course of operations, we assume that operational data management systems will be employed at all times to keep track of physical locations of data. Failure to track media, or refresh data to new media, will of course lead to a loss of the ability to read it at all.

6. Access control

The definition of a standard for Business Acceptable Communications (BAC) presumes the existence of software and services that can use the metadata which is associated with a BAC object. "Resolver" services, for example, are envisioned to translate Terms and Conditions metadata into concrete prices, permissions, and data views. The presumption is that Terms and Conditions metadata will be expressed in abstract categorical terms, rather than in concrete terms. Thus rather than stating in metadata that the retention period of a record is a specific date or a number of years from creation, the sophisticated user will place into the metadata the rule under which disposition should take place, so as to accommodate future change. Similar resolvers can negotiate other access and use metadata to accommodate such changes as inflation in prices, dilution of restrictions based on elapse of time since the transaction, differences in access rules based on specific characteristics of the requester, or re-assessment of risks associated with secrecy or confidentiality. The "resolver" must be put in place by the owner, creator, or manager of a record, and it must be maintained so that it correctly enforces rules. It is presumed that resolver applications will be maintained by those interested in restricting rights. As a practical matter, their on-going operation can be ensured by establishing a socio-legal mechanism that allows users access to the records if, after seeking out the resolver where the record indicates it should be found, they find that no restrictive permissions manager is operating.

Strong pressures are operating in contemporary organizations and society to encourage the encryption of communications. It may be that organizations will continue to feel that communications need to be encrypted, even when they are encapsulated. While resolvers could, in principle, hand users back decryption keys, it would, in my opinion, be extremely dangerous to permit objects being archived for evidence to remain encrypted when written to recordkeeping systems because the encryption method becomes a dependency and is very likely to prevent records from being accessed over time.

7. Retrieval and delivery

Archivists and records managers will not be able to design good retrieval systems, or define the most important retrieval metadata or how to represent it, unless they study the questions which potential users ask.[27] For the time being, archivists will doubtless continue to employ both content-based and provenance-based methods of discovery because they don't adequately understand which, if either, better satisfies users needs.[28]

The metadata required for evidence specifically requires (provides warrant for) documentation of the business or functional context of records creation. If intellectual control is provided based on functional analysis of business transactions and the relationship between such functional "*competences*" and the structural units of organizations, the critical issues for recordkeeping

27 An early example of this is reported in David Bearman, "User Presentation Language in Archives", Archives and Museum Informatics, vol.3#4, p.3-7; see also the report of the NIST study in the summer of 1995 of Nebraskans interests in Federal records.

28 Richard Lytle coined the terms and raised the question in his seminal dissertation. The fact that it still has no answer is clear from on-going discussion, such as Chris Hurley, "Ambient Functions - Abandoned Children to Zoos', Archivaria 40, Fall 1995, p.21-39.

will be how to best represent that knowledge. As is clear from the various models (RAD, ISAD) adopted as standards in recent years, items inherit many of the properties of the provenance that created them. Keeping that data in the item solves numerous problems that have in the past been associated with changes in structures and functions over time. Appropriate ways of representing such knowledge so that structure and function are independent of each other have recently been proposed by Chris Hurley and were explored a decade ago in articles by Richard Szary and myself on multiple independent authority files.[29] It is evident that archivists and records managers will need to maintain such "finding aid systems" to facilitate searches independently of the records themselves, even if the metadata about context is present in the records.

It may also be necessary in the future to search for records that satisfy criteria based on their content, even though this is not essentially a recordkeeping requirement. The Reference Model for BAC is designed to hold metadata that can satisfy such requirements but it is not currently populated by metadata designed to support Networked Information Discovery and retrieval (NIDR) such as that recently proposed by the Coalition for Networked Information and by the U.S. library community.[30] Since the volume of records that are created has always defied cataloging individual records and the content description of records, which are not created to be about their content but rather as a consequence of business transactions, tends in any case to be either misleading

29 David Bearman & Richard Szary, "Beyond Authority Control: Authorities as Reference Files in a Multi-Disciplinary Setting" in Karen Markey ed. Authority Control Symposium (Tucson AZ, ARLIS/NA, 1986) p.69-78 ; Chris Hurley, "The Australian ('Series') System: An Exposition", The Records Continuum op.cit. p.150-172; Chris Hurley, " Problems with Provenance", Archives and Manuscripts, Vol. 23, No. 2, Nov. 1995, p.234-259.

30 See, David Bearman, "Developments in Metadata Frameworks", Archives & Museum Informatics, vol.10#2 p.185-188, and "The Research Process, Metadata and the Image as a Document" in this issue

or inadequate, we may find that these metadata fields are rarely populated.

Ironically, archivists will be able to focus more of their attention on content based access because the provenance based issues can be resolved by implementation of the business process analysis methods required at the time electronic records are created and by the assignment of contextual metadata at the item level. Not only will archivists be able to provide full-text searching of the contents of textually based records, they should be able to augment content-based access by using metadata needed for structural and contextual evidence. For example, we know that the meanings of words in a document are dependent to a great extent on where they appear; each genre of document has different structural elements and the occurrence of a word in one indicates that this was the author, while in another location within the document it references the recipient, the subject, or the object of the transaction. In addition, words having meanings within particular business contexts, and it is quite different if we find a statement that indicates a job was done poorly in a reminder from a colleague, a performance report by a superior, or an audit report by an outside agency. These differences in meanings reflect the fact that words fill out a "frame" (to use artificial intelligence terminology) of discourse in a given domain. Retrieval systems could bring domain knowledge and document genre knowledge into play in content searches of metadata documented objects because the references to business context and document genre are present. This should make the otherwise haphazard aspects of full-text searching much more rigorous. First, domain knowledge can bring the correct thesauri into use in term lookup. Second, domain knowledge and genre knowledge together can suppress meaningless terms and weight meaningful ones in full-text. Third, domain knowledge can provide interpretive frames in which the words used in a particular document can be properly related to each other. Finally, genre knowledge can locate the parts of the record content that should be the source of retrieval terminology.

8. Use history

Recordkeeping systems will store and provide access to metadata-bound evidence.

Sometimes the purposes of such access will be to make use of the data content of records in subsequent business transactions which create their own records. When needed for these purposes, records from recordkeeping systems may be copied to information systems which require their content, but the record itself will never be deleted from, or changed within, the recordkeeping system except with specific records disposition authority. In addition, when copied to an information system, the record will lose its "recordness" and becomes just information, available for incorporation into a new transaction. These transactions will take place through application systems, which like most information systems, are not designed to make or keep records. They will keep data of use to the ongoing work of the organization. Incorporating a record as a record into a transaction (as when records are attached or forwarded to a further authority) is a different business transaction than the original. Engaging in a transaction that incorporates previous records will therefore result in the creation of a new record which will have different business rules associated with its scheduling metadata. Because the creation of records should always result in attaching metadata to them describing the time, place and circumstances of their creation and receipt and their contents subsequent transaction can effect scheduling without any change in the basic system logic and without need for post-hoc, sub-aggregate level re-appraisal by archivists when, for example, a relatively routine record that would have been destroyed in six months, becomes the object of an audit or legal proceeding.

Sometimes the purpose of access is simply to view the records outside of the business purposes of the creating organization. Traditionally such reference uses of archives have not created new records, although logically they are the record of the use of the archives which is itself a function of the organization. In an

evidential environment, viewing a record in conjunction with a business transaction creates a new record for the recordkeeping system and leaves a transaction trail in the original record. This is far more documentary than we have been in the paper age, but the rationale for it is consistent with evidence as well as reflecting realities of the electronic environment. For example, organizations which are succeeding in implementing executive or management decision support systems used to have mid-level managers who wrote reports on issues of corporate strategy. These reports would then go to the executive who would make decisions based on them. Now the same executive can query corporate data bases, display results in graphic form, and make immediate decisions based on these queries without any record being retained of the basis for the decision.

Implementers will recognize that when a user requests a record, a copy of that record is passed to the information retrieval subsystem, but if the user opens the record contents under the control of another application, the contents are incorporated within the application in which he or she is working and while the contents will become part of the contents of a new transaction, the record of the prior transaction will not be. If the user intends to append or forward a record, they will do so in a step that does not involve opening the record. They will need to invoke an applet that in incorporates an encapsulated version of the record within the current transaction.

One of the most exciting aspects of item level control using metadata encapsulated objects is the opportunity it provides for allowing copying of archival records while still monitoring, and if desired, limiting, uses. This permits archivists and records managers to maintain control over records, from the moment of their creation, while permitting distributed custody and enabling distributed access. Users can obtain copies of records, and once cleared to use them by the terms and conditions resident in the object metadata interacting with the resolver, will be able to open

the record, and exploit its informational value, without risk of confusion between the manipulated information and the real record. Multiple convenience copies of this sort can be kept under little or no control without threatening the integrity of the actual record. From the point of view of the active records manager, the fact that all such information is a convenience copy means that it can be destroyed at any time by individuals who don't need it any longer. Implementations could be devised to allow such informational copies to be kept past the period that they would normally be destroyed under retention requirements simply by devising a method that jettisons the record metadata at the point of expiration of the record on which the information is based. This deprives the information of its warrant but ensures that no copies of records, however non-evidential they have become, could survive past the scheduled retention.

In discussions with recordkeepers in Australia in June/July 1996, it was clear that different business rules operate in different organizations to govern whether, and when, use history transactions are written to the originating records. There was agreement that redacted releases, transactions involving filing and indexing, and dispositions under approved schedules all required use history documentation. These actions could be described as recordkeeping transactions of little importance in themselves but considerable impact on the future meaning or uses of the original record.

A second, quite different class of transactions was identified as possibly requiring use history. These transactions were highly consequential in themselves and were considered to be worthy of recording in the original record use history because they reflected on such a significant use. Examples of such transactions were files that were sent to the highest level of the organization (such as the board of directors) for approval or review and/or records involved in lawsuits, commissions of inquiry, and other externally driven discovery processes. Doubtless each organization will devise its own rules about what types of uses to document.

IV. Item-Level Electronic Recordkeeping

1. Technology Challenges

Can organizations implement architectures that remove records from active information management systems at the point of transactions and satisfy the functional requirements for evidence in recordkeeping by capturing and retaining metadata required to reconstruct the context, structure and content of all transactions?

I believe that they not only can do this, and do it for legacy systems as well as new systems, but that the adoption of the Reference Model for BAC within a single organization, together with management of resolvers within that organization, is as practical as universal adoption and does not depend on it. Widespread implementation could come about either by each organization writing and implementing its own methods for capturing metadata and encapsulating objects or by archivists and the business community insisting that software developers and networks implement standards such as the Reference Model for BAC. Since metadata content must either follow an external standard or contain its own declarations (e.g.,. meta-metadata), it would be greatly more efficient for the society at large if instead of requiring individual organizations to implement systems in ways that supported the requirements for evidence, a standard for communications could be adopted that placed the burden for creating metadata encapsulated objects on the application software and network software developers.

However, even if only records conforming to the BAC model were permitted to travel between networks, and even if API and network software developers built all the needed tools for implementing metadata encapsulated object stores, the major tasks facing an organization intent on implementing the model would be the same:

- First, records have to be captured with metadata documenting the nature of the transaction and the business consequences of having a record of this type. This requires up front analysis of business processes and the implementation of information systems with appropriate transaction-stamping both for routine and non-routine applications.
- Second, resolvers need to be maintained to enforce business rules for captured records.
- Third, the authority will need to build finding aid systems containing metadata and content analysis to support queries by both the staff needing access to current records and others desiring access to inactive records at a later date.

These three requirements will have significant impacts on the profession, because they transform what recordkeepers actually do.

2. Professional Challenges

Will recordkeepers be willing to fundamentally transform their business processes and radically change their day to day jobs? Will archivists be able to see their future as managers of virtual records?

Item level control from creation obviates the need to accession, arrange, process, and describe records after "transfer", or to even take custody. These are the processes which have, historically, been the largest part of the archival and records management role and consumed the greatest resources. On the other hand, item level control forces archivists to invest in the front end tasks involving not just scheduling and appraisal but also definition of requirements for description and access. These responsibilities have traditionally been associated with records management, and records managers have by and large not been able to successfully do them. If archivists take on the new tasks, they will find that their role is more one of steering than of doing. And they will come to recognize that they need new skills in systems analysis, business

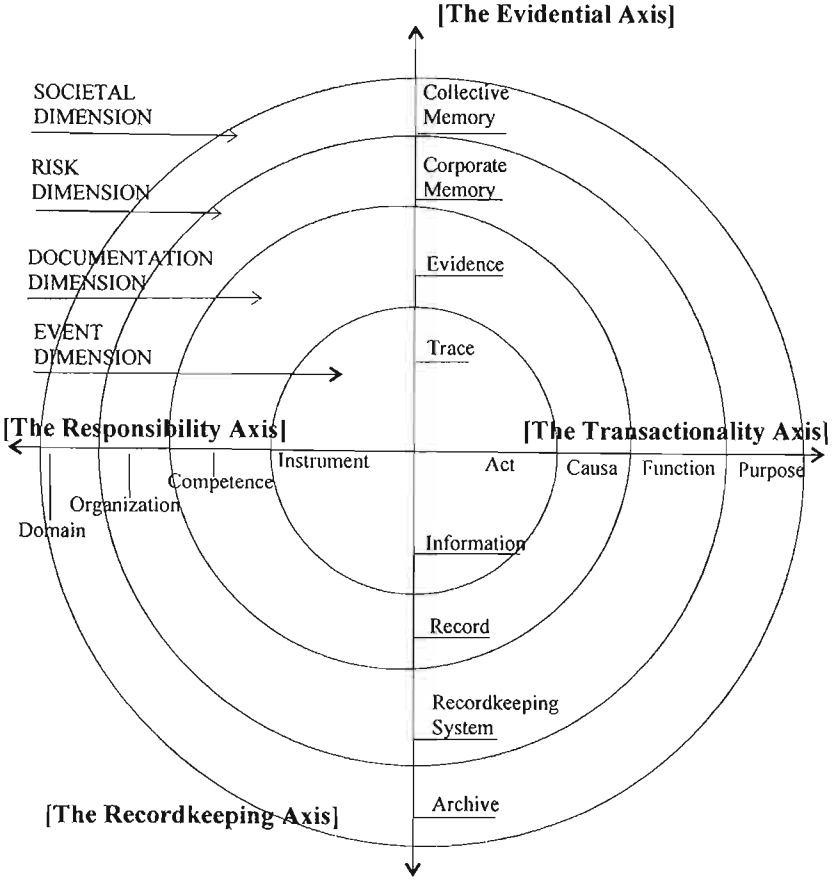
process analysis, and knowledge representation in order to be of assistance to the line managers who will be responsible for and accountable for, good recordkeeping.

Archivists will also need new allies and will need to stop thinking of themselves as "information professionals". Archivists are recordkeeping professionals and that their organizational allies are senior management, lawyers, auditors, Freedom of Information Act and Privacy officers, and citizens who need records to substantiate claims, rather than information professionals. The shift away from the information professions will need to be pronounced in order for people in the organization to understand the rediscovery of records, but it will not make their lives as users of, and dependents on, electronic information systems easier. As recordkeeping professionals, archivists will need to forge peer working relationships with information technology professionals to successfully identify electronic record dependencies and make records migrations work, and to maintain the front-end software that supports the capture of records, but they must take care not to confuse their goals with those of IT professionals.

In the long run, archivists and records managers will find that they can give the rear-end tasks of providing access to librarians and information service centers whose mission it really is and focus their attention on generating the metadata on which successful retrieval will be based. This metadata should support access either by provenance (contextual metadata) or by content (structural metadata and full-text analysis), and should be assigned to items based on front-end practices. Given distributed metadata encapsulated objects and the resolvers with built in access and use rules others can easily provide good information retrieval services to clients without the assistance of recordkeepers.

A framework for this new professional role and philosophy is being articulated by faculty of Monash University in Melbourne Australia where Frank Upward, Sue McKemmish, and their colleagues Chris Hurley and Barbara Reed are proposing to substitute

Records Continuum model as adapted from Frank Upward



a "records continuum" approach for that of the traditional "records life-cycle".[31]

31 See, especially, Frank Upward, "Postcustodial Structural Properties", Archives and Manuscripts, Vol. 24, No. 2, Nov. 1996 (forthcoming) and "The Continuum: Principles, Structures and Dualities",Archives and Manuscripts, Vol. 25, No. 1, May 1997 (forthcoming); Chris Hurley, "Standards, Standardisation and Documentation", paper presented at the Australian

The Records Continuum Model is constructed around the assertion that management of the record is a continuous process from the moment of creation. Management issues arise from four dimensions which are not related to the age of the records but to the point of view of the observer. My presentation of their position here reflects my own use and elaboration on their insights as they were developed during the Monash University sponsored workshops in Melbourne and Canberra in June and July 1996 and is not the orthodox position as represented in their own work, but it shares with them the view that traditional life-cycle frameworks focus professional energy on tasks that archivists and records managers engage in over the life of the record but that these do not add value to the record and make a fundamental distinction between the pre-archival and archival "life" of a record which is misleading. In the continuum model, the record comes into existence at the moment of the transaction and requires continuous care from that time forward until its disposal. It does not pass through phases, although issues in it on-going management can be understood as reflections of its life in four dimensions:

The first dimension, to which I have given the name the Event dimension, consists of the act, the trace, the instrument and the information. In this dimension, the transaction has yet to take place.

The second dimension, to which I give the name the Documentation dimension, is likewise characterized by four attributes along the same dimensions: the act becomes the business transaction or causa; the trace becomes the evidence; the instrument becomes the competence; the data becomes the record. In this dimension

Society of Archivists Conference, Alice Springs, May 1996; Chris Hurley, "Ambient Functions - Abandoned Children to Zoo", op.cit.My graphical representation differs from that used by Frank Upward and ue McKemmish but owes its origins to theirs and to the Monash University workshop in Canberra where it was developed on July 2, 1996.

the act is witnessed by the system and the transaction becomes evidence.

The third dimension, to which I give the name Risk, is characterized by function, corporate memory, organization, and recordkeeping system. In this dimension, the record is appraised by the organization and either kept or destroyed.

The fourth, or societal, dimension has the attributes of purpose, collective memory, domain, and archives. In this dimension the society gives meaning and institutional form to its record.

The attributes are themselves related along spokes or axes which are called the evidential axis (trace, evidence, corporate memory, collective memory), the transactionality axis (act, causa, function, purpose), the responsibility axis (instrument, competence, organization, domain), and the recordkeeping axis (data, record, recordkeeping system, archives). It may be easiest to understand this in a graphical representation, below.

The records continuum is a pedagogical and conceptual framework that can help reunify recordkeeping around its proper focus, the documented event. As such it supports the control of records at the item level as described in this paper and places the issue of control of the record from the moment of its creation within the context of the event that gave rise to the record and the organization or person whose activity it documents. It places the archival and records management tasks conducted by any given organization into the context of the society as a whole and the evidence of an act. Along its various dimensions it informs a novel view of the nature of recordkeeping activity and its social purposes. It seems to me a further example of how the intersection of traditional archival methods and the requirements for electronic recordkeeping has yielded useful and interesting new perspectives on the nature of archives and their management. One of the most important of these, I believe, is the return to item level control.

ARTICLE

The Indiana University Electronic Records Project: Analyzing Functions, Identifying Transactions, and Evaluating Recordkeeping Systems - A Report on Methodology

Philip C. Bantin and Gerald Bernbom[1]

Introduction

Managing and providing access to electronic information has the potential to radically alter the way archivists and records managers do business. It could and likely will transform every archival function, from initial appraisal to long-term preservation. This transformation has had such a profound effect on the design of the Indiana University Electronic Records Project, the authors would like to review briefly recent debate on three very basic questions pertaining to the management of records: 1) How do we define provenance? 2) What is a record? and 3) What is the focus or objective of records management?

How do we define provenance?

Traditional definitions equate provenance with the office of origin, resulting in an appraisal and descriptive system which focuses on administrative structure and organizational setting. But at least since the 1980s efforts have been underway to redefine provenance in terms of the functions and transactions which generated the record.[2] Part of this redefinition has resulted from recognition of the changing nature of modern institutions, characterized by a flattening of the organizational structure and a more decentralized, fluid environment. In this environment, organizational structure and setting have much less relevance in understanding the nature and significance of records than they did in the more traditional, hierarchical business structures.[3] Electronic re-

1 *The authors owe special thanks to Jeannie Kellam, Data Analyst, University Computing Services, Indiana University for her critique and comments, and for her contributions to revision of the methodology described in this paper.*

2 One of the most influential, early articles to stress a functional approach was the article by David Bearman and Richard Lytle, "The Power of the Principle of Provenance,"Archivaria 21 (Winter 1985-86): 14-27. From that same time period see also JoAnne Yates, "Internal Communication Systems in American Business Structures: A Framework to Aid Appraisal," American Archivist, 48 (Spring 1985): 141-58. One of the earliest applications of functional appraisal was by Joan K. Hass, Helen Willa Samuels and Barbara Trippel Simmons, Appraising the Records of Modern Science and Technology: A Guide (Cambridge, Mass: MIT, 1985). This methodology was later applied in a university setting in a publication by Helen Willa Samuels, Varsity Letters. Documenting Modern Colleges and Universities (Metuchen, N.J: The Society of American Archivists and Scarecrow Press, 1992), and to the documentation of high technology companies in Bruce Bruemmer and Sheldon Hochheiser, The High-Technology Company: a Historical Research and Archival Guide (Minneapolis: Charles Babbage Institute, Center for the History of Information Processing, University of Minnesota: distributed by the Society of American Archivists, 1989). For articles advocating a functional approach as it relates to electronic records see especially, Terry Cook, "Electronic Records, Paper Minds: The Revolution in Information Management and Archives in the Post-Custodial and Post-Modernist Era," Archives and Manuscripts 22 (November 1994): 300-328; Margaret Hedstrom, "Descriptive Practices for Electronic Records: Deciding What is Essential and Imagining What is Possible,S Archivaria 36 (Autumn 1993): 53-63; David Bearman, "Diplomatics, Weberian Bureaucracy, and the Management of Electronic Records in Europe and America," in David Bearman, Electronic Evidence. Strategies for Managing Records in Contemporary Organizations (Pittsburgh, PA: Archives and Museum Informatics, 1994): 261-66; and the Australian Archives Home Page at: www.aa.gov.au/AA_www/AA_Issues/KER/KER4.html

3 For summaries of the changing business structure and its implications for recordkeeping see Richard Cox, "Archives and Archivists in the Twenty-First Century: What Will We Become?" Archival Issues 20, no. 1 (1995): 97-113; John McDonald,SManaging Records in the Modern Office: Tam-

cords have strengthened the argument for a redefinition of the concept of provenance. In the world of "second generation electronic records archives,"[4] office of origin becomes far less critical as a source of information about the creation and use of records. David Bearman and Charles Dollar assert that in this electronic environment provenance is more closely linked to the mission, function and transactions of an organization than to organizational units.[5] The implications of this redefinition of provenance around functions and activities are widespread and extend as far as a reexamination of our most basic tenets, beginning with what is a record?

Redefinition of Records

Developing precise definitions or strategies for identifying records traditionally has not been a high priority for the archival and records management communities, but the emergence of electronic records and the ongoing reexamination of archival methodology have changed all this. In the last several years archivists and records managers have begun to put forward explicit strategies to identify electronic records. In line with the redefinition of provenance along functional lines, a number of archivists and record managers are redefining records not as a collection of data or information or as physical objects but rather as the consequence of a business event.[6] More specifically a definition of records provided by David Bearman is gaining widespread support: Records are evidence of business transactions that document organizational functions and provide accountability.[7] We would like to emphasize two points in this definition.

First, archives collect evidence, not simply data or information. And what is evidence? Bearman defines evidence as residing in the conjunction of data (i.e. "the record of the words, numbers, images and sounds actually made by the creator"), structure (i.e. "the relationships among these data as employed by the record creator to convey meaning") and the context (i.e. "the relationship

ing the Wild Frontier,S Archivaria 39 (March 1995); 70-79; David Bearman and Margaret Hedstrom, "Reinventing Archives for Electronic Records: Alternate Service Delivery Options," in Electronic Records Management Program Strategies, 82-98. Ed. Margaret Hedstrom, (Archives and Informatics Technical Report No. 18, Pittsburgh, PA., Archives and Museum Informatics, 1993); Michael Hammer and James Champy, Reengineering the Corporation: A Manifesto for Business Revolution (New York: Harper Business, 1993); Thomas H. Davenport, Process Innovation. Reengineering Work through Information Technology (Boston, MA: Harvard Business School Press, 1993); and David Bearman, "Diplomatics, Weberian Bureaucracy, and the Management of Electronic Records in Europe and America," in Electronic Evidence, 254-277.

4 This term was used by Terry Cook in his article "Easy to Byte, Harder to Chew: The Second Generation of Electronic Records Archives," Archivaria 33 (Winter 1991-92): 202-216.

5 David Bearman, "Recordkeeping Systems," in Electronic Evidence, 43-44, and Charles Dollar, Archival Theory and Information Technologies. The Impact of Information Technologies on Archival Principles and Methods (Macerata, Italy: University of Macerata, 1992), 48-51.

6 For descriptions of the evolution of the concept of the record and redefinitions of the term see Richard Cox, "The Record: Is It Evolving?" The Records and Retrieval Report 10, No. 3 (1994): 1-16; Richard Cox, "The Record in the Information Age: A Progress Report on Research," The Records and Retrieval Report, No. 1 (January 1996): 1-16; David Roberts, "Defining Electronic Records, Documents and Data," Archives and Manuscripts 22, No. 1 (May 1994): 14-26; Glenda Ackland, "Managing the Record Rather Than the Relic," Archives and Manuscripts 20, No. 1 (1992):57-63; David Bearman, "Managing Electronic Mail," in Electronic Evidence, 188-191; David Bearman, "New Models for Management of Electronic Records," in Electronic Evidence, 283-84; Charles Dollar, Archival Theory and Information Technologies, 45-48; and the Australian Archives Home Page, Archival Issues, "Keeping Electronic Records," at the URL listed above.

7 This definition can be found in several of Bearman's writings but see especially, "Archival Principles and the Electronic Office," in Electronic Evidence, 147.

between the record and the activity out of which it arose").[8] If any of these attributes is missing, the result is data or a non-record.

And second, archives are records of business transactions. A record is not just a collection of data but the consequence of a transaction. This redefinition has tremendous consequences for archival management. It shifts the focus of our work from asking what kinds of records we should keep and from a primary concern for content analysis to an emphasis on the concept of how the record was created and used.

Goal of Records Management

Traditionally the focus of any program was on managing records throughout their life cycle. This may not be a sound strategy for any records management program, but particularly in an electronic environment. If we are refocusing our sights on the transaction producing the record rather than the actual record itself, it makes much more sense to focus records management not on the records but on managing the recordkeeping systems throughout their life cycle. If the systems which capture, maintain and support retrieval of records can be demonstrated to be sound, it is argued, the records within that system will be sound.[9]

These redefinitions of provenance, records and the objectives of records management are central to the goals of the IU project. In the project we focus on functional analysis as the means to understanding records creation; we define a record as the evidence of the transaction, with evidence defined in terms of content,

8 David Bearman, "Archival Principles and the Electronic Office," in Electronic Evidence, 148.

9 For descriptions of the recordkeeping system concept see David Bearman, "Recordkeeping Systems," in David Bearman, Electronic Evidence, 34-70, and the Australian Archives Home Page, Archival Issues, "Keeping Electronic Records" at the URL listed above

context and structural information; and finally, our objective is on analyzing the recordkeeping system rather than the record itself.

Description of the Indiana University Electronic Records Project

In June, 1995, IU began a two year project funded by the National Historical Publications and Records Commission (NHPRC) and Indiana University to implement functional analysis methodology and to test the ideas regarding functional requirements for recordkeeping systems and the critical role of metadata put forward by David Bearman, Richard Cox, and the project personnel associated with the University of Pittsburgh Electronic Records Project.[10] As a test site for these ideas, project personnel at IU would be attempting to determine what worked and what did not in the context of a large university system, and what, if any, additions or revisions might be made to the functional requirements and metadata models developed by the University of Pittsburgh project. Project personnel identified four stages of development for the project.

Stage 1: Functional analysis of business units, identification of business transactions, and identification of the basic information categories necessary for establishing evidence of specific business transactions.

Stage 2: Identify and describe existing recordkeeping systems (or information systems which perform recordkeeping functions), for the previously-identified business transactions.

Stage 3: Evaluate existing systems in terms of the "Functional Requirements for Evidence in Recordkeeping" and the "Metadata

10 For the most up-to-date list of the functional and metadata requirements see the University of Pittsburgh Electronic Records Project Home page at http://www.lis.pitt.edu/~nhprc/

Specifications Derived from the Functional Requirements" developed at the University of Pittsburgh.

Stage 4: Develop implementation tactics. Basic goals here are to review modifications to record systems in terms of the tactics - policies, standards, designs, etc. - best suited to ensure implementation. In determining when and if to implement, costs, risk of not implementing and organizational priorities will be considered.

And as a final step in the two-year project: Critique and, if necessary, recommend revisions or additions to the "Functional Requirements for Evidence in Recordkeeping" and the "Metadata Specifications." And, more broadly, critique the methodology developed in the IU project, including its use of the Pittsburgh models, in terms of cost, time required, user acceptance, and ability to influence the implementation of recommended system improvements.

Most of the work in the first year has focused on Stage 1, although we have made some progress on Stages 2, 3, and 4 as well. In this article, we will be concentrating on Stage 1 activities, i.e. on the development of the methodology designed to collect the data needed to answer the questions posed in the University of Pittsburgh "Functional Requirements for Evidence in Recordkeeping" document. Developing a methodology for practical implementation of the "Functional Requirements for Evidence..." was a time-consuming and sometimes difficult process. Especially for the tasks in Stage 1, it was necessary to define the specific steps in this process; there was no existing methodology to draw upon. There was some excellent information on functional decomposition in the literature of the data administration field.[11] But this methodology says very little on identifying transactions. Information on the methodology used in information systems analysis and

11 An excellent source on this subject is James Martin, Information Engineering (Englewood Cliffs, N.J.: Prentice Hall, 1989).

design is plentiful and useful, but this literature tells us little or nothing about creating and maintaining records as evidence of a transaction.[12] And what about archival and records management literature? We regret to report that the literature in these fields was not very helpful in developing the methodology. Archivists and records managers are not yet focusing on these issues, and regrettably, most are not trained to perform these functions. After some difficulty and a few false starts, we have arrived at the approach presented here. We will continue to refine this methodology over the course of the next year or so. Nonetheless, we think the steps outlined here will provide us with the information we need to review the information systems according to the requirements outlined in the University of Pittsburgh study.[13]

Methodology Steps: IU Electronic Records Project

1.1 Functional Analysis
    1.2 Identify Business Transactions
    1.3 Identify Evidence of Transactions: Data, Context, Structure

2.0 Review Existing Recordkeeping Systems/Information Systems

3.0 Evaluate Existing Systems Using the "Functional Requirements..."

4.0 Recommendations and Implementation Tactics[14]

12 An outstanding source on this topic is Edward Yourdon, Modern Structured Analysis (Englewood Cliffs, N.J.: Yourdon Press, 1989).

13 To view all the documents produced by our Project, consult our Electronic Records Project Home Page at http://www.indiana.edu/~libarche/index.html.

14 Examples, like this, used throughout our paper should be understood as provisional products of our work. The methodology we are using is still being evaluated and developed.

Methodology (1.1): Functional Analysis

In performing functional decomposition, we initially addressed three basic questions: what is a business function? how does one gain the information needed to do the analysis? and what are the products of the process?

This analysis will not succeed unless one has a very clear understanding of the concept of a business function, and one knows and can articulate how a function relates to business processes, transactions and activities. Functions, in fact all the concepts named above, relate to official actions pertaining to the business or mission of the enterprise. Using the distinction drawn by Thomas Davenport, the focus of our analysis is on business functions, not business processes.[15] In our project we broadly differentiate a business function from a business process by emphasizing that a function describes WHAT is done in the organization; this is in contrast to a business process with its emphasis on outputs for a particular customer or market and on HOW work is done. Functions (and sub-functions) are comprised of one or more related transactions, which are themselves comprised of related activities.

What follows are examples of business functions we have identified thus far in our project in the business areas of Financial Management Services and Student Services.

Examples of Business Functions:

Financial Management Services

- Generate Accounting Transactions
- Manage Capital Assets

15 For definitions and examples of these concepts see Thomas H. Davenport, Process Innovation, 5-9, and the list of definitions provided on the Indiana University Electronic Records Home Page at the URL listed above.

- Create and Maintain a Chart of Accounts
- Administer Contract or Grant
- Pay Employee Compensation and Taxes
- Pay Non-Employee Taxes
- Purchasing and Maintaining Accounts Payable

Student Services

- Collect Fees and Maintain Student Financial Records
- Maintain Academic Records

We found that documentation identifying functions and sub-functions existed within our target offices. We relied on two basic sources: system documentation and interviews with one or more staff from the business area. The major objective then became one of verifying and refining the data, usually by adding sub-functions and narrative information about the functions. We do not yet have enough experience to say whether archivists can routinely expect information about functions to be readily available, but we can say with some confidence that data managers readily understand this concept of functional decomposition, or are already familiar with it. In identifying people to interview regarding the business functions, we suggest starting with those individuals who understand the entire business function. At this point in the analysis, one needs an overview of the business area, not details of individual activities. This usually means talking with senior-level staff and managers from a business area, rather than the individuals who mange the data systems on a day-to-day basis.

The products of this analysis consist of a list of first-level functions, subdivided into more detailed second-level and, if necessary, third-level sub-functions. Based on prior experiences with functional decomposition within the university, our general guidelines were that even major business areas typically have only six to twelve first-level functions, and that for each of these first-level functions, there are typically between three and eight

second-level sub-functions, with the lower numbers being very common. We found, however, that to fully understand the nature of the function, project staff had to go beyond a simple listing of functions and sub-functions. What had to be developed were short narrative descriptions of what occurred within the given function. These descriptions were designed to answer the following questions: who is involved in executing the function, what is the nature of the action, how do the primary activities within the function relate to one another, and what are the products of the function?

What follows is an excerpt from a functional analysis completed by project personnel for the function - Maintain Academic Records.

Example of Functional Narrative:

3.0 Maintain Academic Records
3.1 Evaluate and record transfer credits

Input is transcript from another college or university. Evaluate in terms of equivalency. Some equivalents have become routine, others require judgement. Outputs include evaluatorUs notes and paper document (may be called "transfer credit evaluation form"). Data from this document is entered into online system. The studentUs electronic file, the Historical Course File, is updated. Journal file is updated with information about the update: what data, updated when, by whom. Audit trail is produced from the journal file and this report is compared to source documents to verify accuracy. Paper record and/or source documents are microfilmed.

3.2 Record student courses completed (with grades)
3.2.1Record original grades

Receive records from sub-function 2.6 (course registration lists). Sort by selected criteria. Produce paper roster which is also mirrored electronically. Distribute paper roster to faculty. Faculty

assign grades and return to Registrar. Data entry process takes place. Paper rosters of updated files are printed and compared to the original source documents for validation. Data from electronic roster file is re-sorted back into records-by-students, and the Historical Course File is updated. Identifiable records are produced by semester. (etc.)

Methodology (1.2): Identification of Transactions

The process of identifying business transactions proved to be a more difficult activity than functional decomposition. The identification of business transactions is a relatively new concept for archivists and the IT community alike. The absence of practical precedence manifested itself at the most basic level - in defining just what we were looking for. Our first step was to develop a working definition of a transaction with which the project staff was comfortable and which we could articulate to our audiences. The working definition adopted by the project is as follows. A transaction is: 1) an official action, related to the business or mission of the university; 2) a public, not private, action involving more than one person; 3) an action undertaken, having a beginning; 4) an action completed, having a definite endpoint. These are considered necessary, but not sufficient, criteria for defining a transaction. They help in distinguishing between transactions and non-transactions (e.g., actions that are not public, not official, not completed).

But even with these criteria, considerable judgment is used in identifying transactions for the purpose of this analysis. We view the transaction as a conceptual tool used to organize our description of business activities in a way that discloses their meaning. This is not to say that the transaction is an arbitrary unit created for the sake of the analysis. The concept of the transaction is clearly grounded in actual and real functions and activities within a business area, but the size and boundaries of the transaction we choose to document is viewed as a product of choice and judg-

ment. In practice, this means that we seek to combine as many small related activities or transactions as practical into a larger transactional unit, while ensuring that no function or activity goes totally undocumented.

In moving from functions to transactions, the project staff considered the following: 1) What is the official action? 2) Who is taking or initiating the action? 3) What objects are being acted upon? and 4) What individuals are interacted with? We found that a transaction narrative statement consisting of a single sentence which identifies the individual or agency who is acting (subject-phrase), the official action (verb-phrase), the individuals or objects who are acted on or interacted with (object-phrases), and any clarification or further specification of the action, actor, or objects of action would provide us with the information to address these questions and to identify transactions.

What follows is an excerpt from the set of transactions for the function - Maintain Academic Records.

Example: Maintain Academic Records: Transactions

Transactions - Overview

1.0 University awards academic credit to a student.

2.0 University awards academic degrees to a student.

3.0 Faculty member assigns a grade to a student for a course.

4.0 University evaluates college/university transcript for transfer of credit.

5.0 University certifies completion of academic credits and degrees to external agencies.

Transactions - Detail

1.0 University awards academic credit to a student.

1.1 Transfer Credit. University awards academic credit to a student based on courses taken at another college or university.

1.2 "Regular IU" Credit. University awards academic credit to a student based on completion of "regular IU courses" (courses offered and enrolled in during all or part of a standard academic term - semester or summer).

1.3 "Independent Study" Credit. University awards academic credit to a student based on completion of "IU independent study courses" (courses offered by and enrolled in through the School of Continuing Studies, not associated with a standard academic term).

1.4 "Other" (or "Special") Credit. University awards academic credit to a student based on a test or examination, or other demonstration of proficiency. (etc.)

Methodology (1.3): Identifying Evidence of Business Transactions: Data, Context, Structure

According to our definition, establishing "recordness" requires identifying the transaction and preserving the evidence (data, context and structure) of that transaction. In the previous steps, we identified the transactions. At this point, we determined that a step was needed to identify the basic information types or categories required to provide evidence of these transactions. Knowledge of these information types will be essential when we attempt to analyze the systems in terms of the "Functional Requirements." And we felt that inserting this step at this point in the methodology made more sense than trying to gather this information at some later point.

Identifying categories of information is not the same as examining the data themselves. First, our focus at this point is on a conceptual model of the information required to document a

business transaction, and not on the computerized records actually stored in an existing information system. Second, our focus is on the abstraction of information-types rather than specific instances of stored data and data-values. From the beginning, our goal has been to determine through analysis of business functions and transactions what kind of information is needed as evidence. This is similar to the process that system designers might have gone through, but with a different emphasis. While system designers might emphasize the use of information to satisfy requirements of business operations or management decision-making, our focus is on information as it satisfies requirements of accountability and evidence. But like a systems designer engaged in requirements analysis and definition, our primary question was: What information do we need - in our case to describe and document these business transactions - and not, what information do we happen to have?

To answer this question, methods of conceptual modeling, including conceptual data modeling,[16] are adapted and applied to this task. Based on the narratives of business functions and transactions, and supplemented by interviews with business area staff, we identify categories of information that represent evidence of the transaction. These information-types are found by addressing such questions as:

1) What action was taken; who took the action? (Data)
2) Within what domain (e.g., temporal, organizational) did this transaction occur? (Context)
3) What relationships exist among data associated with this transaction? (Structure)

16 A practical introduction to the topic of data modeling is found in C.C. Fleming and B. Von Halle, "An Overview of Logical Data Modeling," Data Resource Management (Winter 1990): 5-15; a more thorough coverage is found in R.G. Ross, Entity Modeling: Techniques and Applications (Boston, Mass: Database Research Group, Inc., 1988).

These and other questions like them help us focus the analysis on information that is relevant to the "Functional Requirements for Evidence in Recordkeeping" and is specific to the business-relevant facts of a particular transaction.

Regarding the "Functional Requirements," it is our observation that the Pittsburgh Requirements do not entirely address the issues involving the "completeness" of a record. There is a requirement that records be "Complete." This requirement is comprised of three sub-requirements: "Accurate" (quality is controlled at input), "Understandable" (relationship of information items and their representations supports their meaning), and "Meaningful" (contextual linkages carry information necessary to understand the transaction).[17] Based on our analysis to date, we identify a fourth element of this requirement, which might be named "Sufficient." Namely, a requirement that the record must contain the information needed to represent the business-relevant facts about the transaction - the official action, the actors, and the objects acted upon - and to uniquely identify each.

What follows is a sample of the information categories identified by project personnel as evidence for a single transaction.

17 See Functional Requirement Number 7 "Complete" on the University of Pittsburgh Electronic Records Project Home Page at the URL listed above

Example: Information Categories Necessary for Evidence of the Transaction:"University awards academic credit to a student."

Transaction:

1.1 Transfer Credit. University awards academic credit to a student based on courses taken at another college or university.

* About each student
  Information to identify the student
* About each college the student attended
  College where courses were taken
  Dates of attendance at that college
* About each transfer credit awarded
  Course identifier: Department, Course number
  Credit hours
* About each awarding of transfer credit
  When credit was awarded
  University office who awarded credit
* About relationships among information-types
  Each student may have attended many colleges (or one, or none)
  From each college there may be many courses (or one, but not none)

Methodology (2.0): Review Existing Record Systems

At this stage, the staff now attempts to describe how existing information systems actually collect and manage data. The first task in the process is to identify the information systems currently used to support the business transactions we have identified. This information would typically have been gathered as part of one of the earlier interviews with staff in the business units. The process

of review is based on existing documentation (technical, procedural, and policy-level) and on interviews with staff in the business units and in the computing organization. The principle product of this step is an organized collection of documentation and field notes which describe: the primary data stores and data items in the system, primary processes or processing cycles, overall flow or movement of data, primary users of the system and its data, and procedures or policies governing the operation and management of the system. In gathering this documentation and writing field notes, the project staff always keep in mind the questions posed by the Pittsburgh "Functional Requirements." This has the effect of better defining the search for documentation by focusing the review on those items relevant to evaluating the system as a recordkeeping system.

It is only when all this data is collected that the staff can begin to address the questions posed in the Pittsburgh "Functional Requirements" document. We have now identified 1) what we are trying to document (the transaction), 2) the basic information categories necessary for establishing evidence of specific transactions, and 3) what existing systems, if any, currently manage this information. We are now ready to determine to what degree our system is capturing and retaining records, i.e. evidence (comprising information on content, context and structure) of a transaction.

Implementation Issues

Given the preliminary nature of our analysis to date, we have not yet recommended or negotiated for any changes in the information systems we have reviewed and analyzed. But project personnel have discussed some basic concepts in regard to implementation, and have initial thoughts. First, we recognize, as do the individuals associated with the University of Pittsburgh Project, that institutions will not likely implement everything in the "Functional Requirements" document, nor should they. The decision to implement requirements will be based on a variety of factors,

including costs and benefits, risk of retaining or disposing of documentation, and organizational needs and priorities. But, we agree with Bearman when he writes: "Decisions not to satisfy functional requirements are just that; they do not invalidate the requirement."[18]

Second, we are not certain whether we can justify broad implementation of this methodology for the large number of primary transactions performed by the University. To date, the time and cost of analysis has been high, and consequently, it may be necessary to limit the scope of our analysis to the primary and high priority functions and transactions as defined by a team of University personnel. On the other hand, much of the time and cost can be attributed to the learning process and the effort required to develop and refine the methodology. In the coming months we will have more accurate data about the time and cost required, minus the burden of development and learning.

Third, we need to address an over-arching question behind our work: How effectively can an existing information system satisfy the requirements of a recordkeeping system? We see two separate approaches to the use of computer systems for preserving evidence of business transactions. One strategy is to maintain the records within the present information system. The other strategy is to create a separate recordkeeping system apart from the information system.[19] At the present time, our plan is to evaluate the feasibility of the first strategy, and we hope that our work will give guidance on the question of whether a business information system can also function as a recordkeeping system. To implement this plan we will have to address several major problems. One challenge will be to find the means of capturing records within systems that are also designed to collect data or non-records. Another major issue will be to develop a means of preserving inviolate the records of these transactions within an active, ever-changing information system. In short, a major challenge will be to integrate recordkeeping requirements into data management practices.

Finally, there is a major implementation issue which involves identifying the mechanisms for coupling documentation or metadata to the data itself. Again, there are two basic strategies. On the one hand, one can encapsulate or bind the metadata to the data. In this arrangement, the two components (data and metadata) move together as a unit.[20] The other strategy is to link the data to the metadata by means of direct links to computerized metadata or pointers to non-computerized documentation. This issue relates strongly to the previous one. Computer systems whose inherent design specifically addresses the requirements of recordkeeping are better able and more likely to bind data and metadata; existing information systems will more probably depend on links and pointers to external metadata and non-computerized documentation.

Concluding Remarks

What have we learned about the methodology? We would like to highlight four points.

1) We have learned that analysis of functions and transactions is a more accurate and useful indicator of record creation and use than the examination of organizational structure.

18 David Bearman, "Recordkeeping Systems," in Electronic Evidence, 59.

19 For a description of a project designed to create a separate recordkeeping module within a system see Mark Giguere, "Philadelphia Electronic Records Project-Phase I Update," The Philadelphia Record 1, No. 2 (September 1995):9-10, and the Philadelphia Electronic Records Project Home Page at http://www.phila.gov/city/departments/erms/erm.html.

20 For a description of a project designed to create "metadata encapsulated objects," see citations on the Philadelphia Electronic Records Project listed above.

2) We have learned that much data gathering and analysis must be completed before the Pittsburgh "Functional Requirements" can be applied, and that this process is time consuming and potentially expensive. However, we have spent much time in developing and refining the methodology. With a methodology in place, we will now get a much better reading on the time and costs required to complete this process.

3) We have learned that a key to the success of the methodology is the identification of the transaction to document. A business area might conduct dozens or even hundreds of transactions which fit the definition we have established, but it is unlikely that we would choose to fully document all of these. To do so could be cost prohibitive, and our experience has shown that many small transactions can be grouped together to form a single, larger transactional unit. The challenge becomes one of determining at which point, or at what level of generality the documentation process will occur. It may also be that in some business areas or even for the institution as a whole that the best strategy will be to create a priority list of transactions to be documented, based on generally agreed upon criteria.

4) We have learned that identification of functions and transactions, and the analysis of systems, demands a strict methodology and a set of skills that most archivists and record managers were never taught. We have discovered that some of the skills and methodologies employed by the IT community, particularly data modeling and systems analysis, are invaluable to archivists. But we have also learned that the IT community does not have all the answers. The lesson here is that interdisciplinary collaboration is essential. We have discovered that neither an archivist nor a technologist alone can adequately design and implement this methodology. But our experience would indicate that together they possess the knowledge and skills necessary to make it work.

Museums for the New Millennium, September 6-7, 1996

This conference, held at the Smithsonian Institution to mark its 150th anniversary, brought an extraordinary number of senior museum personnel from throughout the world to Washington to explore issues confronting museums in the future, and in a manner consistent with that future, its proceedings were made available to remote audiences monitoring the conference over the World Wide Web with full text transcripts within hours of the conclusion of each session. I attended virtually, as you can still at: <mignon.si.edu/proceed1.htm.>

Secretary Michael Heyman opened the conference identifying private fund raising and dealing with public controversy as the two major challenges facing his institution. He was followed by Jennifer Jarrett a futurist who identified globalization, metropolitanization, changes in middle class values, images as the primary means of communication, life-long learning and a knowledge-driven world as the major trends leading to the year 2025 - but failed to make very meaningful connections between them and the challenges facing museums.

Harold Skramstad, the first speaker in the next session, on Changing Public Expectations of Museums, argued that museums, having become professional, now need to turn to how they can better serve their potential clientele. So what? - he asked - as do many of our visitors. Future uniqueness of museums as institutions, he suggested, will be based less on their collections than on their claim to authority, their connectedness and their trustworthiness. Iran Hiring of the Japanese American Museum in Los Angeles followed with a discussion of how a museum designed

to interpret and celebrate one culture can reach out to others, in part through joint programming. Neither the talks nor the discussions addressed the stated issue of public expectations, but all agreed that leadership in museums in the future needed to be more open to and able to orchestrate public opinion.

Ellsworth Brown spoke in place of Nina Archibal in the next session. He suggested that museums were already well on their way to the future and that their power is in the connections they make and have made in the lives of visitors. "Culture", he said, "is a way of locating yourself in a world." This led him to explain the phenomenal growth of science centers. But to me, Brown's recent discovery of the concept of knowledge representation, from reading the Research Agenda for Networked Cultural Heritage, was the most fascinating insight. I hadn't realized that museum directors don't ask the question of how their representations of knowledge map to the perspectives of the customers.

The next session was dubbed Town Meeting. It reported on break out sessions that had been taking place during the day. Major ideas emerging from these meetings were that museums were generators of culture, not just collectors/reflectors and the role of the museum as a forum for the community, a place of story-telling and engagement. In general the groups reflected a shift in museum thinking from collections to audiences. It was agreed that collections remained the special province of museums (though they might also be collections of ideas) but that online interaction would play an increasingly important role. The term communicating, and especially communicating with communities, recurred regularly. The problems museums faced were identified as the need to broaden (from a very narrow base) their audiences, to exploit technology, and to integrate the business culture with the museum culture without being overcome with commercialism.

In the evening, both actual and virtual visitors toured a technology showcase. For those of us who live with these things every

day, it was pretty tame, but I expect that the on-site audience found more of interest.

The second day of the conference opened with a session on transforming the internal structure of museums. Bob Janes, CEO of the Glenbow museum in Calgary whose major point was that to make organizational change requires courage and persistence, in addition to sound strategies. It also requires true teamwork within the organization and the confidence of professionals in themselves which allows teamwork to thrive. This confidence extends to welcoming ambiguity, change, and risk.

A session of international perspectives, which followed, brought museum directors from Russia, Mexico, Australia, Brazil, and India together to reflect on the museums of their futures. Issues raised included the increasing costs of presenting museum exhibits, illicit traffic in cultural treasures, the potential of the popular, non-professional museum in a community, the needs of ethnic groups and aboriginal peoples to find identity and define appropriate uses of culture, the potential of friends organizations, and partnering.

In the final session of the conference, Bran Ferren, Executive VP for Creative Technology R&D at Walt Disney Imagineering spoke to the "Future of Museums - Asking the Right Questions". Ferren argued that the World Wide Web, not as it is today, but as a broad band interactive communication medium, is the most important technology in the history of mankind because it carries language (the first great invention) and storytelling (the second) and writing, pictograms of course (the third), publishing (the fourth). Museums, he suggested need to establish context to deliver great content and to establish community (especially over the Internet) by customizing a hybrid experience consisting of virtual and real.

Betsy Broun, Director of the National Museum of American Art wrapped up by talking about how to tell stories with objects,

while referencing the Vietnam Memorial (a deconstructionist statement), the Olympics (an abstraction with political consequences), and Judy Garlands slippers from the Wizard of Oz (concrete objects related to a myth represented in movies). All these and other cases she saw as instances of connections made about people as individuals. The challenges she saw were buying the necessary talent and understanding the technology and its implications.

In the end, what was the Conference. An event that tells more about the past of museums, I think, than of their future. The speakers suggested that the birthday party of a museum should connect it with its community, make its objects live for others, use the technology to interact. In the event, they organized a party for the professional elite. There was much talk of objects and stories but none were constructed. The community was not present nor reached. It is interesting to think what kind of event would have occurred if the speakers had been the planning committee - would they have taken the risks Bob Janes suggested and launch the story Bran Ferren imagineered?

CNI/OCLC Workshop on Metadata for Networked Images

In the last issue of this journal I reported on the background to a meeting held in Dublin Ohio September 24-25 on metadata required for access to networked images. Over sixty invited participants struggled throughout the two day meeting to define both what metadata was required for images and how the metadata requirements fit into the previous work on (text) documents from the 1995 Dublin conference and the proposed Warwick framework for a metadata architecture. In the end the participants arrived at two conclusion which are not surprising in themselves, but which are very surprising outcomes for a group gathered to

define the difference between image and text requirements for metadata.

First, we concluded that the metadata about documents which are images is not very different from the metadata about documents which are texts (and quite helpfully established that there are non-document like objects, both textual databases and graphical ones, like GIS and image generation systems, to which the "document-like" metadata does not apply).

Second, and very gratifying to me personally, we established that the necessary metadata is congruent with the metadata required for evidence as described in the Reference Model for Business Acceptable Communications which resulted from research I conducted last year <www.lis.pitt.edu/~nhprc/meta96.html.>

I was also excited that the meeting adopted the "stages of the research process" approach to determining what metadata required in which packets that I had proposed as a method for parsing elements into packets. My notes indicate that the working meeting concluded that:

1. Discovery is the first stage in the research process, which we can provisionally describe as including: Discovery, Retrieval, Collation, Analysis, and Re-presentation
2. The requirements for each stage are distinct (though not yet well specified). Different metadata will be needed to satisfy the requirements of each stage so these ought to be the criteria by which contents of metadata packages are defined.
3. Descriptive metadata (that which documents the item-in-hand and the original-item of which is a representation) will be quite different for fixed/bounded/document-like objects and for dynamic or non-document-like objects.

4. Document-like objects include texts, images, movies, musical performances, speeches and other information objects which are characterized by being fixed (e.g., having identical content for each user). Non-document-like objects include such information as virtual experiences, databases (including ones that generate document-like outputs), business graphics, CAD/CAM or geographic information generated from database values, and interactives which might have different content for each user. In the context of image discovery, these sources do not "contain" images as much as they "generate" images. The images they generate may be described as fixed document-like objects, but the metadata required to describe them (the systems doing the generating) are distinct.

5. Common discovery requirements for document-like objects include the ability to search for items by:
a) Identification metadata:

   The unique id, handle and/or persistent identifier of this item or collection.

b) Instance or Fixation metadata:

   Who created this image or digital file and when, who published it/when?

c) Source image metadata:

   Who created the original content, when was the image captured?

d) Content metadata:

   What, when and where does the image depict? (pre-iconographic description). Some content metadata may be calculated automatically.

e) Subject metadata:

   What is the subject matter of the image? What is its genre and object-type?

f) Context metadata:

   Why were these images created, in what business process and by whom?

g) Structure metadata:

   What is the file structure, encoding, compression and format?

h) Relationships metadata:

   If more than one item, how are they related to each other. For any single item, what are its relationships to other data (including non-image data)? How are the described objects related to intellectual schema's of the user's discipline?

i) Terms & Conditions metadata:

   Are there restrictions on access and use? [To answer definitively, the system may need to acquire data about the user or proposed uses]

j) Use history metadata:

   Where has this item been previously published, referenced, used?

6. The discovery process returns retrieval metadata packages:

   Location/ Identification metadata: (handle)

   Terms & Conditions Metadata: Access/Use resolver requirements

   Structure metadata: Physical (File) documentation & software/hardware dependencies

7. Retrieval returns collation/analysis/representation metadata packages:

Context metadata: Full provenance schemas

Content metadata: discipline specific schemas

Structure metadata: data specific local storage schemas being supplied

Terms & Conditions Execution metadata (redaction reporting)

8. Collation, Analysis and Representation return use history metadata package:

9. Some uses, typically including publication and redacted release, would typically be reported to the use history metadata. (Use history metadata is generated according to the business rules of organizations making or documenting these uses and differ in different business environments. Libraries often record none, special libraries often record publication history, archives often record publication and research use history, medical records centers usually log all uses.)

At the conclusion of the meeting, I presented what I felt was a useful way of looking at the metadata and stages and preparing ourselves for a way to think about the architecture of a system that would support this process. Some work is going ahead along these lines.

Metadata Packets and Research Stages

Possible Metadata Elements in each stage as contributed by each packet:

Description					
Instance	Publisher Name /Publication place & date	Order number	Ordinals of relationships		
Source	Creator, Date of creation, object-type, title, genre				
Content	Coverage Date/ Location; Color	Arrangement ; file schema's	Data Content Dictionary	Data Values Dictionary	Conventions of representation
Subject	Topic/theme				
Context	Functional Provenance	Agents & Roles	Rules	Value Tables & Organizational Schemes	
Terms& Conditions	Access and Use Resolvers	Agreed terms/ proposed uses	Acknowledgement	Anonymization requirements	Credit line
Structure	File type; Size, format	Resolution, Compression method, Dependencies			
Relationships	Item/Collection/Site	Disciplinary schema/ class			
Use History	Publication/ release citation	Captions, citation details			

The final report of the Dublin Image Metadata Workshop has not been released yet. When it is completed, it will doubtless be published widely. I think the process stands a very good chance

of influencing Internet metadata practices and should be followed carefully by those in the cultural heritage arena. A few points made at the meeting should be kept in mind in assessing these and other metadata solutions because to make a system that will come as close as possible to consistently and comprehensively identifying sources on the Internet will require us to adopt some common ground rules.

- The target objects must include objects which are not in digital format but for which metadata is available in digital format (eg. the Workshop n Metadata for Networked Images became the Workshop on Networked Metadata for Images)
- The target objects must be largely self-describing if they are in digital format on the network (and they can be because much of their data is metadata for discovery)
- Conventions in broad use for documenting original objects should be extended to citations of those objects as the sources for digital objects (and applied by documentalists already in the library cataloging tradition)
- The system must accommodate metadata made by a variety of agents, for different reasons, at different times in the life of the object (and the model for this should support orthogonal metadata, perhaps with registered objects and data sets, although practice mat see overlaps).
- To be used effectively elements of metadata must be readily available as required by each stage in the research process in which the user is engaged (though different implementations might deliver some metadata at stages prior to its being needed).
- The user needs to know the elements of metadata (or at least categories) available but should be shielded from implementation protocols

Museum Computer Network, Ottawa October 30-Nov.2, 1996

For the third year running, MCN has been fortunate enough to have an international meeting as its conference partner - this year it met in Ottawa with the Canadian Heritage Information Network. Not only did this assure an additional hundred or more Canadian participants, it attracted dozens of museum professionals from around the world. The result was a conference of five concurrent sessions and three days of exhibitions. With this embarrassment of riches it is no surprise that participants were simultaneously elated and overwhelmed.

Actually, there are other reasons to feel these conflicting emotions. Just as the computer has made its way to center stage in the museum, its character has changed to that of a communications device. The technical staff which has been supporting implementation of systems is suddenly thrust into the role of being the public face of the museum, its publishing arm, content generators and the focus of public attention. These are not roles for which they are suited nor will playing these roles further the real interests of museum computing over time. But the WWW has brought them to exhilarating new heights.

While many of the sessions dealt with traditional concerns such as vocabulary control, database management, integration of in-house systems, and implementing commercial collections management, the real topic throughout was getting the museum online, delivering an experience to remote visitors, and programming for interactivity and the public. Unfortunately, because these topics were often not the actual subjects the level of discussion of Web related issues tended to be fairly general.

Several sessions excited me because they showed the Internet in the way that it can be used by museums. In one, Susan Hazan of the Israel Museum in Jerusalem explained how they had col-

lected depiction's of Jerusalem from children around the world, and used the Internet to link the children to their displayed works. Then they installed C-U-See-Me to allow two way interactive teleconferencing for remote visitors to see the galleries and talk with on-site visitors. Sarah Kenderdine explored how her museum in Western Australia has used the Web to create a virtual library of underwater archaeological digs taking place around the world under various auspices and how the visitors can examine the materials uncovered in these digs, test them and conduct research online.

In another session, speakers from Canada's CultureNet, the New York Foundation for the Arts ArtsNet and the Getty Information institute inspired Los Angeles CultureNet all explored the ways in which creating a framework for communal participation led to the generation of culture and the interaction of culture creators with audiences. Each speaker was able to explain how these cultural networks different from traditional, authority laden, central publishing and the reasons why the communities they serve responded to them.

In each case the Web enabled museums and cultural institutions to take on new roles. They were not just online brochures, or collections management databases being pushed to the public, they were communications experiences in opening up ways of reaching new audiences. This, indeed, is the promise of museum computing. If the technical staffs can persuade themselves to take on roles as internal consultants they can enable the curators, museum shop managers, museum educators, and scientists to reach new publics and establish the museum as a critical part of the broader community. If they try to be the content providers, ultimately they will occupy an overlooked corner of their institutions or will be perceived to have failed and will have their brief taken away.

CALENDAR

Feb. 8-12

Cannes, France, Milia'97. see *www.reedmidem.milia.com*

February 11-15

New York, Visual Resources Association, [Registration: Arleen Arzigian, Boston U., Art History Dept., 725 Commonwealth Ave., Boston,MA 02215]

March 16-19

Los Angeles, Museums and the Web. *www.archimuse.com\MW97toc* or, contact: Archives & Museum Informatics, 5501 Walnut St., Suite 203, Pittsburgh PA 15232 USA; phone +1-412-683-9775; fax +1-412-683-7366

Call for Papers:

ICHIM'97, Le Louvre, September 1-5, 1997

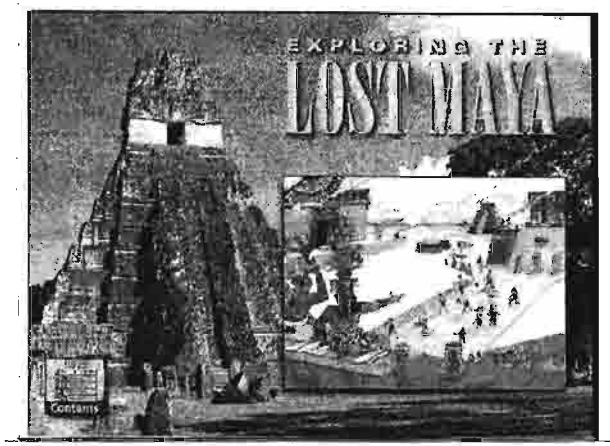
The Fourth International Conference on Hypermedia and Interactivity in Museums invites proposals for papers, sessions, or demonstrations for the Paris occurence of this bi-annual event. Abstracts (and papers) may be submitted in French or English; simultaneous translation will be provided at the conference. For the full call for participation and program as it evolves, see: *www.archimuse.com/ichim97* and *www.louvre.fr/ichim97*

Send proposals to: David Bearman, Conference Organizer, 5501 Walnut St. Suite 203, Pittsburgh PA 15232 USA or (preferably) to: *dbear@archimuse.com*

New Forms of Multimedia Expression in Exploring the LOST MAYA CD-ROM

Ecaterina Geber

Exploring the LOST MAYA is the second[1] in a series of CD-ROM reference works on the world's ancient civilizations from Sumeria[2], the San Francisco based multimedia publisher. This volume covers the Maya world and civilization of Belize, Guatemala, Honduras and Mexico, spanning the period from 2000 BC to 1700 AD and linking it to the present situation by showing the current state of development and reconstruction.



The title presents historical sites, exploring the evolution, culture and demise of the ancient Mayan people. Interactive maps are gateways to the site details, image index, travel information, Quick Time VR site panoramas, historical narrative texts[3] and historical photos. The user, exploring the Lost Maya, is provided with material in previously unimaginable abundance, and is given considerable control in handling the content. Just as feelings have been captured adequately in literature and in film, **Exploring the LOST MAYA** shows that they will be conveyed by interactive multimedia products in near future. **Exploring the LOST MAYA** is, apart from everything else, a new form of communication.

The collection of information encompasses a quantity and variety that traditional information media could hardly manage:

- 700 images of 37 sites in four countries;

1 Preceded by the award winning **Exploring Ancient Cities** - an exploration of four ancient civilizations: Petra, Pompeii, Teotihuacánand four of Crete's Bronze Age palaces - all images, text and movies being exportable and reusable for non-commercial applications.

2 **Sumeria** - Bryant Street, Suit 3D, San Francisco, CA 94107,USA, Tel. 415.904.0800, Fax. 415.904.0888

3 By **Robert J. Sharer**, Professor of Anthropology and curator of the American Section of the University Museum of Archaeology and Anthropology at the University of Pennsylvania, the editor of the Fifth Edition of the Ancient Maya (1994), Stanford University Press.

- 95 events along a timeline starting with the appearance of villages in 2000 BC and ending with the fall of Tayasal - the last Maya capital in 1700 AD;
- video, narrated and written text on Temple Construction, Calendars & Astrology, Writing Systems and Ritual Life including kingship, bloodletting, sacrifice, ballgame, trade and burial;
- 148 images of material culture objects;
- movies and 93 slides on early explorers as Stephen & Catherwood or Alfred Maudslay.

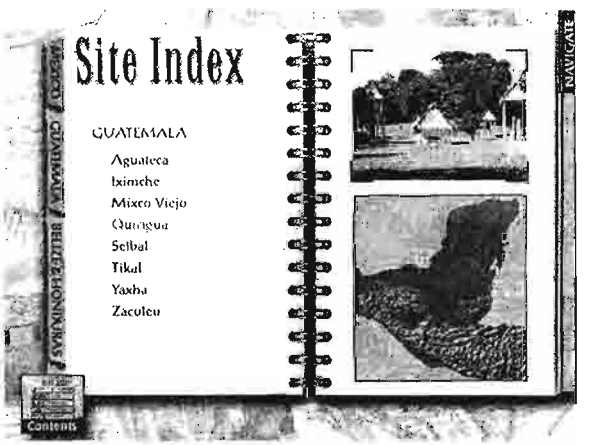
One of the reviewers, I asked for comments, admitted that he valued being able to choose what was personally important, move much more easily, quickly and focused using the CD than any comparably large maze of data involving text, pictures, voice and video.

The CD structure, imposing connections that arouse specific associations and experiences, turned out to be more accessible than a book. The **accessibility of the structure** is based upon the balanced conceptual representation.[4] The core of this conceptual representation - **Contents** - is formed of six coordinated image and text bars (navigation buttons) combined in a way which builds a flexible classification judged to be consistent

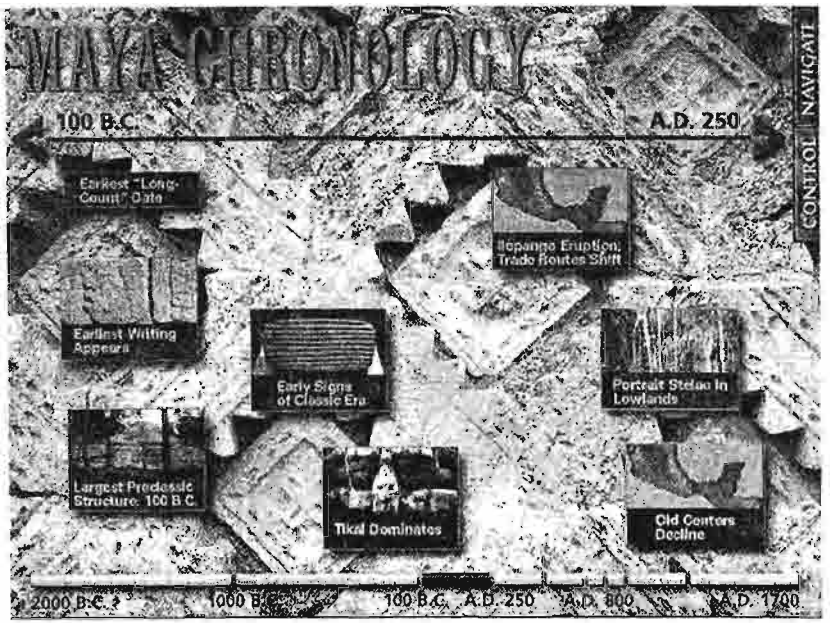
4 The classificatory, analytical and symbolical processes.

members of the same class. The classification processes operates at all the levels of the resource and relates pieces of content to each other in terms of 'a kind of' or 'part-whole', modelling structured, goal oriented exploration. A click on **Maps** displays an interactive map with a number of hot sites, a click on a selected site, brings forward a site map with several hot buildings, a click on the building reveals a zoomed image with a text explanation. Each map incorporates the six image and text bars: Image Index, Slide Show, Getting There, Panorama, History & Description, and Historical Photos.

The **Site Index** (below) indexes sites both spatially and alphabetically.



Chronology is an analytical timeline. **Ancient Culture** refers to writing, ritual and cosmology. **Material Culture** classifies artefacts. The **Early Explorers** chapter presents photos and movies from the nineteenth century.

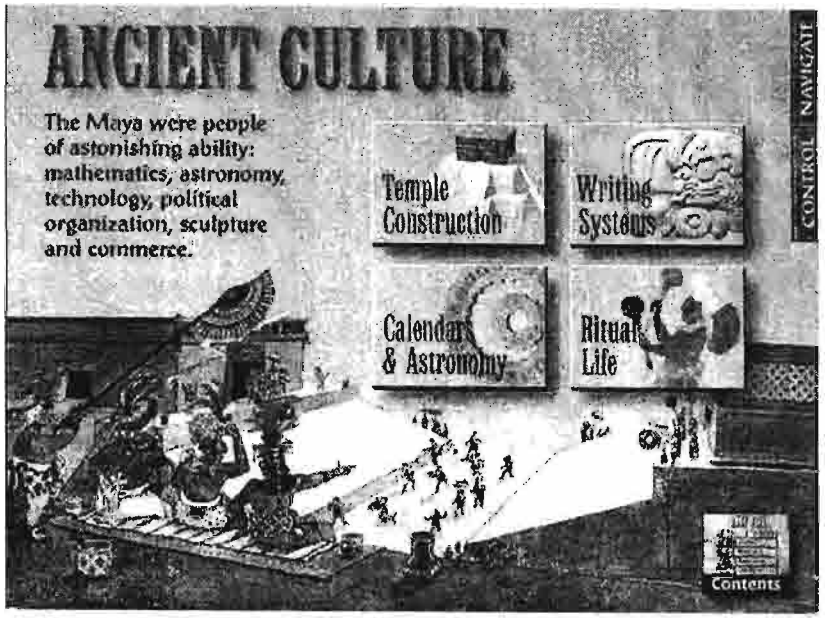


The reader may notice that this structure does not permit rapid to a set destination. After each vertical step, there is a horizontal, choice. The developers managed to provide a model for exploration built on consistent options which helps the user to overcome the feeling of being abandoned or lost, a feeling often experienced by CD-ROM adventurers.

The special attention awarded to the navigation rules is reflected in the fact that the navigation buttons are mirrored by a pull-out menu or a sliding bar, permanently on the right of the screen, inviting to action, under the title: **NAVIGATE**. The only objection that might arise is that in some situations, a click on the background takes the user back while in others it does not.

The strength of **Exploring the LOST MAYA** comes from the way in which it succeeds in transmitting the feeling that the Maya world is the land where man, nature and time have converged for ever. While it is meant to be a documentation medium, it has an

effectiveness all of its own in presenting existing material using an attractive and new manner of presentation.



The **LOST MAYA** land becomes a virtual playful context where the parameters of interaction become known through the process of playing with the system building up both fictional and documentary aspects.

This computer controlled resource becomes a site for investigation where the image offers an encompassing view of a subject

enacted through participation. It is precisely as a place where passivity of observation is superseded by the necessity of action that one can grasp a new understanding form, which is both generative and informative. A new understanding of subjectivity appears as well, one that accounts for both the image and the behaviour it initiates. In coupling the image with its performance, enriched by a conversation and a dialogue structure, the act of exploration implicies more than it reproduces. Within the flow of images, there is a set of possibilities, some controllable, some reorganizable in a subjective manner.

Reoganization patterns reuse the material in a stimulating way building new contexts. The same image, for example, can be found in different contexts: in the introduction, during the spatial investigation (MAPS), under the timeline procedures (CHRONOLOGY) or among the artefacts of sites and collections (MATERIAL CULTURE). Being integrated in new contexts, the repetitions have the effect of creating familiarity rather than redundancy.

Ultimately, the **LOST MAYA** CD-ROM achieves this multimedia expression quality through the richness and variety of

stylistic means. It contains films, slide-shows, virtual reality scenes, 3-dimensional views, interactive maps, graphic animation, simulation, composed images, photographs, drawings, stories, facts, tutorials, nature, artefacts, archaisms, neologisms, music, sounds of nature, function and navigation buttons, linearity, interactivity, and a complex mix of oral, written and hyper-texts.

In my enthusiasm to grasp the multimedia aspects, I should not forget to mention however that there is still a long way to go. It is

obvious that the richness of the content engaged the developers' attention. Some of my reviewers pointed out that there is a discrepancy between the very careful treatment of some parts and

visible haste in others. Probably that is why the labels are very often placed as if there was a time limit. Watching the slide shows, one is very much impressed by their coherence and the interactive or linear choice offered to the user, but it is hard not to notice the visually intrusive labels. There was also some graphic tension between background and foreground information. One kept looking for signals such as those on the site maps when exploring interactive zones but no change occurs.

The GETTING THERE section is useful but not very inviting. I wondered if the information provided or the maps offered would be helpful if I was really going there, because after having the chance to explore the LOST MAYA on disc, I definitely wanted to visit in person. Similarly, my reviewers, when exploring the **ANCIENT CULTURE** section, asked for more information, more levels, and more interaction, which suggests a limitation even in this stimulating and educational resource.

One final remark. One of my users wanted to compare historical photos with current ones so I was delighted to be able to point

out the generous facility offered by the publisher which allows users to **Print Screen**, **Export Text**, **Export Photo** and **Export Movie**, usng a sliding **CONTROLS** bar (always present on the screen) which enabled my user to take this comparison as far as he wished.

Those who are interested in additional information, might find useful to address the Mesoamerican WWW Page presenting extensively resources, news & lists, software, news & info, pictures, Precolumbian link page.**Exploring the LOST MAYA** is available on the Internet CD-ROM shop, priced at $30.98.

INBOX

Reports & Proceedings:

* Joint Information Steering Committee, "JISC Five year Strategy: 1996-2001", July 1996

Following on from recent papers on exploiting information systems in higher education <info.mcc.ac.uk/NTI/JISC/Jisc-Issues.html> this major steering body within the Higher Education Funding Councils of the UK issued a five year plan (available free in paper from the JISC at Northavon House, Coldharbour Lane, Bristol BS16-1QD ENGLAND, and at <www.niss.ac.uk/education/jisc/strategy.html>. The document is interesting in part because it tries to cover all the territory, from infrastructure through content which will be required for a future national higher education information service. In 243 numbered paragraphs it ranges from the most abstract principles to the most concrete actions, never quite convincing nor quite failing to sustain interest. There is much grist here for anyone thinking about technology futures, but to my mind no blueprint or even a clear trajectory.

* National Archives of Canada, "Preliminary Study of the Core Competencies of Future Records Specialists", May 1996.

Price Waterhouse was commissioned to develop a preliminary list of core competencies from interviews with archives managers and with key information management personnel through the Canadian government. More in-depth focus group sessions with users, functional experts from corporate services (legal and financial) and the information management/information technology communities led to the conclusion that two levels of records specialists would be needed in the future. The senior records specialist would be university trained and expected to serve as systems designer, policy driver, retrieval expert, and advisor/coach. These four roles in-

volved the following responsibilities:

I. Systems Designer:

A. Understands the business of the organization and its processes and workflows as well as the accountability needs of the organization.

B. Understands what types of records need to be captured in the organization.

C. Ensures that record keeping requirements are built into the design of business applications, work processes and management functions.

II. Policy Driver:

A. Develops the rules of record keeping (i.e. policies for what needs to be captured based on the business needs of the organization; procedures; standards, practices for identification, description, classification, organization, storage, protection, retention and disposition.

III. Retrieval Expert:

A. Provides access to information stored in records, and the tools for navigating through information sources in order to locate specific records for users

IV. Advisor/Coach:

A. Raises awareness of senior managers and other users about the importance of records as an information resource, as well as an instrument to help meet accountability requirements.

B. Provides advice and guidance to users on their record keeping responsibilities and how to meet these requirements

The full report breaks each of these job functions down by main tasks and identifies the skills, knowledge and abilities which are required to perform the main tasks. The National Archives hopes after further review of these future competency requirements by the profession that it will be able to influence universities and government agencies to institute programs to educate professionals with these competencies. Current work is integrating the requirements into an RFP for support of government training needs for information professionals and into efforts underway in the Department of Defense to classify job skills. [For further information, contact John McDonald, Information Management Standards and Practices Division, NAC; 613-947-1510]

* Proceedings of the Forum on Technology-Based Intellectual Property Management: Electronic Commerce for Content, edited by Brian Kahin and Kate Arms, Interactive Multimedia Association Intellectual Property Proceedings vol.2, August 1996

The twenty articles in this volume represent the most definitive, technical and broad-based documentation of the state of the art in the technologies for management of intellectual property which have yet been written. All the major stakeholders - the Copyright Office, Imprimatur, Copyright Clearance Center, the AAP, CNRI, IMA, IBM, the credit card companies - are represented, and each used the occasion to present a detailed model and state-of-the-art report to a well informed and critical audience.

Books and Articles:

* Bonnie MacEwan and Mira Geffner. "The Committee on Institutional Cooperation Electronic Journals Collection (CIC-EJC): A New Model for Library Management of Scholarly Journals Published on the Internet" The Public-Access Computer Systems Review 7(4)(1996) <info.lib.uh.edu/pr/v7/n4/mace7n4.html>

Every day I see traditional library operations which happen to take place with electronic resources described as archives. This admirable joint collection development project, designed to acquire electronic journals for academic researchers, is described by the authors as "this archival process" which "will be reviewed by the CIC Task Force on Preservation and Digital Technology, in concert with CICNet computer specialists, in order to identify an appropriate long-term solution. The archive is intended to serve as a perma-

nent record that can be used in the event the original site is destroyed or discontinued, and, therefore, it is not available as a "public site." This model is consistent with traditional archival efforts, in which, for instance, original microform copies are stored, with public access provided through copies."

Want to know what's in the library? "The CIC-EJC has a descriptive Web home page that provides: (1) subject, title, and keyword searching; (2) the ability to view titles in the collection listed either alphabetically or by subject; (3) and bibliographic records for all titles in the collection, including links to the originating journal servers" - what was once called a card catalog, I think.

* Heather MacNeil, "Implications of the UBC Research Results for Archival Description in General and the Canadian Rules for Archival Description in Particular", Archivi & Computer: Automazione e Beni Culturali v.6 #3-4, 1996 p.239-246

This short article presents the results of collective thinking at the University of British Columbia to explain why their decomposition model of archives management does not reflect "levels" which were an important concept in the RAD model and why they feel that archivists will always need to describe records after their receipt in archives rather than using the approach of metadata captured at the time of records creation. The explanation is lucid, although I believe she is wrong. This and other detailed descriptions of the UBC project results appearing in this Italian journal over the past two years should be mandatory reading.

* John R. Smith and Shih-Fu Chang, "Searching for Images and Videos on the World-Wide Web", Columbia University Center for Telecommunications Research, Technical Report #459-96-25 <www.itnm.columbia.edu /webseek/paper/ws.html>

The authors explore a variety of means whereby searches for images on the WWW can take advantage of the fact that images are hyperlinked into documents at relevant points and that text at the link anchors and nearby is marked up in ways that often indicate its relation to the images. The automatic methods developed by the authors produced impressive search results which with further refinement point to some advantages of searching for images in hyperlinked environments.

* Deirdre Stam, "Shared Access to Visual Images - The Potential of the Web", VRA Bulletin, vol.23 #2, 1996 p.63-67

In this short note, Deirdre Stam recalls the original purposes Charles Cutter had in mind in inventing modern cataloging and suggests how a combination of archival control and item-level description can make visual collections more available in a networked environment.

* Eric W. Stein and Vladimir Zwass, "Actualizing Organizational Memory with Information Systems", Information Systems Research, vol.6#2, 1995 p.85-117

The authors explore the concept of an Organizational Memory Information System (OMIS) and how it can be implemented and sustained, exploring its utility to work group and corporate processes and the ways in which technologies such as Lotus Notes and email relate to the technical and social issues raised by an OMIS. Their findings, as with Steins other recent works (International Journal of Information Management vol.15#2) have special relevance to archivists struggling with the need to establish systems architectures capable of ensuring that users create records with contextual metadata.

* Richard Wiggins, "The Mysterious Disappearance of the White House Speech Archive: A Pioneering Application of Technology Vanishes", First Monday, issue #2 <www.firstmonday.dk/issues/issue2/white house>

After reviewing the exciting implementation of indexed sound recordings at the White-

house Web site, the author reports on the disappearance of the speech archive during the 1996 Presidential campaign (presumably because the Whitehouse was nervous about its use by the opposition).The author then uses this as a basis for arguing against agency control over records, seemingly without reflecting carefully enough on the other records under agency control or the technologies that might be available to ensure continued access to records of continuing value. Still, an interesting story and one that could exercise the best imaginings of students trying to architect a solution that would prevent a relapse.

Journals & Newsletters:

* AMIA Newsletter Special Section: Preservation (#33, Summer 1996, p.15-30

This pull out contains articles by Peter Brothers, P.J.Heller, Jim Lindner, Ross Lipman, Douglas Nishimura, and Ken Weissman on the full range of cleaning, storing and digitizing issues in video preservation today.

* Journal of the American Society for Information Science, vol.47 #11, November 1996

This special issue, largely devoted to "Perspectives on Distance Independent Education, is edited by Howard Besser and Stacey Donahue. It contains a range of articles that must be of interest on one of the hottest issues and biggest challenges of the day.

Possible Contributions of the Reference Model of Metadata Required for Evidence to a Reference Model of Metadata Required for Image Description[1]

David Bearman

For the past several years, I have been engaged in defining the requirements for evidence in electronic recordkeeping. One of the results of that research has been a proposal for a reference model of metadata required for evidence. While at first glance these issues may seem incidental to the this workshop, I believe they are not and hope to be able to convey a few points that will assist us in defining metadata requirements for image discovery in networked environments.

My points fall into four major categories:

1) the role of images as evidence

2) the nature of the research process and the requirements for metadata at each of its stages

3) the nature of image resources which exist importantly both as aggregates (as well as individual images) and as non-digital resources (as well as digital ones)

4) the relationship between real users and the architectures and protocols that will be required to carry out the metadata dialog.

1 Presented at the CNI/OCLC Workshop on Metadata for Networked Images, Dublin Ohio; September 24, 1996 For the final report of this workshop, see Stuart Weibel's article in the January issue of D-Lib (forthcoming).

I. Images as evidence and what we found about the metadata requirements of evidence

While many reasons for seeking images relate entirely to their visual content and evocative qualities, a large proportion of those seeking images are looking for documentation of an event captured by fixing it on film, videotape, or digital medium. It will therefore be valuable for us to try to accommodate the requirements for evidence.[2] These requirements indicate that metadata that serves several different purposes and could be created and maintained by several different agents are required to manage records (evidence of trans-actions which are bounded by having been communicated at a single moment). These metadata include some categories not identified in the original "Dublin Core" description elements for document-like objects (by which we believe the authors essentially meant published text).

Specifically, we found that six layers of metadata were required for records to be "Business Acceptable Communications" or evidence. These 'layers' are equivalent to the 'packets' of the Warwick Framework, but because records are a sub-set of documents for which the metadata needed for evidence must always be created at the same time as the transaction recorded, we envisioned them as being encapsulated with the record. These layers are:

1) Identification
   1a) The Record Declaration
   1b) The Handle
   1c) The Discovery Metadata
2) Terms & Conditions
3) Context
4) Structure
5) Content
6) Use History

2 See www.sis.pitt.edu/~nhprc for complete reports on this project and in particular see papers by Bearman and Sochats, "Metadata Requirements for Evidence", and Bearman, "Item Level Control" for discussion of these and other points about the "Reference Model for Business Acceptable Communications" (also on the site) and its implementations.

The record declaration, context, structure and use history layers/packets are not separately identified in the original Dublin Core, so let me comment briefly on their applicability to images:

The record declaration is simply a flag which declares what follows to be an object of a special sort, a record, which by being opened will lose its recordness. It is, in effect, a seal, which can be manufactured by the recordkeeping system copying a record under controlled circumstances but cannot be "put back on" a record after it has been opened. If images had such a flag, we would know that they had not been electronically altered subsequent to their creation.

Context is often thought of as provenance, but archivists have a more complex understanding of the documentation requirements of provenance than is provided by the name of the organization and unit from whom the resource was obtained (the Dublin Core suggested publisher, other agent, date of issuance would be adequate). Especially in the electronic environment, documentation of provenance requires identification of the business process which created the records. Business process are typically a more powerful descriptor of the nature of records than subject indexing. Thus I can describe 500,000 images created by the Minnesota Department of Highways and Motor Vehicles to document the signage on state highways by describing the process of taking photographs every 100 yards from the right and left side of a vehicle along every State highway and you know virtually everything you need to know about those images. Similarly when I describe the business processes generating huge collections of images made in tax assessment, insurance appraisal, environ-

mental surveillance, driver licensing, etc., simply by referencing these processes and you know a great deal about the images in these datasets (including the metadata that is likely to be generated to support their business use).

Structural metadata has special significance to archivists because we need to think about how to keep data longer than many generations of computer software and hardware, and beyond the life of specific computing paradigms. Bits are software and hardware dependent on many levels from their physical encoding, compressing, chunking, and sectoring to their complex intellectual signification as fields, data values, relationships and genres. When we consider the numbers of ways that images can be encoded (bit-maps, vectors, fractals) and what is represented (stills, motion, black & white, color, etc.) the potential complexity of structural metadata should be clear. What is less clear is what structural data will be used initially to qualify our searches and how that can best be represented. I'm fairly sure that the metadata envisioned by the Dublin Core, which was object-type (or genre), form (actually format) and language will not tell us what we need to know for images.

Finally, archivists are accustomed to searches for records that were used in a particular court or administrative case, referred to in a decision process, cited in a brief, etc. and these questions are common in image searches as well for many of the same reasons. Users frequently seek images that appeared in print in a given newspaper or were in the background in a film sequence, using this citation to use history as their search input.

II. The Nature of the Research Process

The metadata workshop in Warwick this spring made significant progress in placing the "Dublin Core" metadata into a framework when they identified the existence of separate packets of metadata that might be created by different agents at different

times. Like the Reference Model for Business Acceptable Communication (evidence), the Warwick Framework assumes that agents other than the document creator could create and maintain metadata. Maintenance may involve dynamic "resolvers" in the network environment which ensure the continued correctness of metadata pointers, as is most evident for persistent resource identification or for maintaining terms and conditions of access and use. I am convinced that one important issue for us will also be to identify when metadata in the image model is fixed "in" the record and when it will be 'resolved' at the end of dynamic pointers.

In formulating the BAC model, we envisioned the researcher encountering these layers sequentially in the course of a reference/use process, but our model was relatively primitive because we only envisioned objects whose metadata was encapsulated with their content and which therefore would communicate all their relevant metadata on the way to obtaining access to that content. This modeled one aspect of the requirements - what the receiving system would need to know to make use of the content and what the content provider would need to know to determine what content to provide and how to convey it.

Since the publication of the Warwick Framework, I've become convinced that a second organizing principle is involved. The research process itself has (at least) five stages - discovery, retrieval, collation, analysis, re-presentation - each of which requires different metadata elements. But the metadata required at each stage will not come from just one package, it will come from many, if not all.

This workshop needs focus its attention on the functional requirements for discovery, and doing so may help us to define the kinds of criteria we need to determine whether a particular element of metadata needs to be, or does not need to be, in the "image core".

III. The Nature of Image Resources

To discover resources we need to be very clear about what resources our metadata are about. Most importantly, we need to distinguish between items and aggregates. The majority of image resources will only be described as aggregates both because functional provenance works (as my story from the Minnesota Historical Society suggests) and because they were acquired and managed as collections and resources do not permit item level description. There are some different things which need to be said about aggregates (which cannot be described as individual files are).

Secondly, we need to decide in this workshop if we are limiting ourselves to digital objects (and especially to ones on the Internet) or if we are actually trying to provide for discovery, using networked digital data, which might discover images that are in analog formats or off-line. I hope it is the latter, because many image resources will not be available in digital formats for a very long time, but information about such resources could easily be provided in the network.

Finally, in the long history of image description we encounter a critical distinction between what the image if **of** and what the image is **about**. I need not remind this group of the importance of Panofsky's distinctions, except to make the obvious observation that the original Dublin Core categories do not make these crucial distinctions and that image object or content (pre-iconographic) will be as important to many researchers as image subject or topic.

IV. Management Issues of Metadata Architecture

Many of the issues that will need to be dealt with after this workshop agrees on what metadata is required for image discovery are properly out of scope here (in particular those having to do with defining syntax for metadata elements and the definition of discrete protocols), but I believe there are a class of architectural issues which cannot be completely overlooked.

I've already mentioned two of the most important:

1) that the metadata we are identifying belongs to different packets which are in turn used at different stages in the research process.

2) that the metadata we need may be the actual values required or it may be a pointer to a dynamic resolver in the environment which will give us back the actual values required (e.g. it might be hard, or soft data). This is particularly important in the arena of terms and conditions for access and use which will vary with who the individual requesting access is and what uses they wish to make and which may vary with the age of the resource, the history of its prior use, and many other variables. The same principle applies to handles which will need to point to resolvers or risk becoming brittle.

Let me mention a third because it could otherwise be a stumbling block and a fourth which I hope we will leave to experiment:

3) I do not think that users should need to know where the metadata they require is held, how it is referenced or how it is internally organized in order to discover a resource, even though each of these kinds of metadata will be crucial to their system at a later stage in the process of using the images. If possible, users seeking resources should be presented only with those descriptive elements they require to discover the objects and the discovery process itself should return metadata that will be required for subsequent steps including location, citation, and the disciplinary and system schema's that control the organization of the items within the aggregates discovered. In this way the architecture can support the special requirements of different disciplines and more detailed uses of retrieved records without burdening our model

with the need to carry metadata not required until later in the research process. This will be an important for extensibility and syntax independence.

4) The Warwick Framework assumed complete orthogonality of the metadata in different packets. I expect that we will find that agents constructing distinct packets often have to, or want to, incorporate overlapping metadata elements. Probably we will need to operate differently on metadata items depending on the source or attribution of the metadata. I hope this can be left to experimentation because it introduces complexity into the model which doesn't help us at this stage and, in any case, cannot be dictated in the ultimate implementations.

VRA Core Categories

In the fall issue of the Bulletin of the Visual Resources Association (vol.23,#3 p.57-63), Dustin Wees, Chairman of the VRA standards committee reports the draft Core Categories for Visual Resources adopted by his committee and invites comments through January 15, 1997 [*by post - Clark Art Institute, Box 8, Williamstown, MA 01267; tel: +1-413-458-9545; fax: +1-413-458-2336; email: dwees@clark.williams.edu*]

The discussion and presentation follow much of the framework set out in the Getty Information Institute's Categories for Description of Works of Art and the 21 Core Categories include only one non-CDWA field (site, for architectural objects). There are 13 categories related to the Object (overlapping about 50% with the 13 categories of the original 'Dublin Core'), plus three categories about creator, and five "surrogate" categories including view description, image type, image owner, owner number, and source.

As Wees states, in his introduction, a standard isn't standard until widely used. One of the inherent problems with this standard is the lack of incentive for institutions to follow it. The slide library community, which is the fundamental maker and potential user of this standard, hasn't been the locus of much information interchange. Unless the standard is adopted as part of an online distribution documentation set (a position more likely to be occupied by the results of the Dublin image standards workshop - see entry below), or as part of a rights information management data set (a position more likely to be occupied by AMICO or some other MESL follow-on entity), following the core will not be of very much value to visual resources repositories. This is not to suggest that the VRA work was not good - just to point out that it is unlikely to be successful unless it hitches itself to another train.

CD Technology Developments

This fall, the long awaited DVD-ROM has been delayed for about six months, but the even longer awaited CD-Erasable and CD-Recordable formats have become available and moved into use. At this stage it is anyone's guess which of these formats will survive into the future. All might coexist since each occupies a slightly different niche or none might survive because consumers and software aren't really ready yet. The good news on all three fronts however is that broad coalitions of drive and media manufacturers have agreed and are building on the ISO 9660 logical format combined with UDF (Universal Disk Format) and its variants UDF-Plus, Micro-UDF, to ensure as much as possible that the three media will each be interoperable with formats of the same produced by other manufacturers, and even upgradable to be read in future drives. The packet writing approaches taken for CD-Recordable (write once) and CD-Erasable (write many) media may prove problematic, but since the expected use for CD-E is backups and temporary movements of large quantities of data, this may not prove a barrier to adoption. What is more likely is that software for the major uses of CD-E will lag behind the medium availability and that the DVD and ultimately DVD writable markets will strangle the CD writables and erasables in their infancy. [*Hugh Bennett provides a useful perspective on these developments in CD-ROM Professional, September 1996, pp.29-44*]

Digital Object Identifier System for Publishing

On September 9, the Association of American Publishers (AAP) announced selection of the team that will develop a Digital Object Identifier (DOI) system for use by the publishing industry. The team, comprising R.R. Bowker, a division of Reed Elsevier, Inc., and the Corporation for National Research Initiatives (CNRI), was chosen following a competitive bidding process initiated last spring as part of AAP's all-out effort to promote development of systems for managing copyright in the digital environment. Copyright management is seen as the key to successful commercial use of the Internet by the publishing industry.

The DOI project is the outgrowth of a year-long AAP initiative to identify the needs of the publishing industry to facilitate safe and successful commercial ventures on the Internet and in other networked environments. AAP's initial research revealed a fundamental need for a unique, unambiguous way to identify digital materials — a type of "electronic license plate" for a "digital vehicle" traveling the information superhighway. The Digital Object Identifier system will serve that purpose. The Bowker/CNRI team will focus on three key areas during the first year: developing a numbering system for identifying digital objects created by publishers, creating an agency for assigning publisher numbers, and developing a network-based directory to link digital objects to their publisher. DOIs will be made accessible via a high-speed computer system developed by CNRI — a scaleable, distributed system on the Internet, with open interfaces, allowing information about digital material to be retrieved almost instantaneously. R.R. Bowker will establish the agency for making publisher numbers and other related information available to any and all publishers. Since the DOI system uses open standards, publishers and other companies can build their own products and services based around DOIs.

[*For further information, DOI project contact at Bowker is Maureen Adamson, Vice President, Business Development—(908) 665-285 madamson@reedref.com.; DOI project contact at CNRI is Constance McLindon, Director, System Deployment—(703) 620-899 mclindon@cnri.reston.va.us.*]

Internet II Architecture

The Internet II Project architecture created at a workshop in Ann Arbor, Michigan, in July 1996 is intended to be a work-

ing document and will be updated as the Internet II project (a collaborative effort among a number of universities, federal R&D agencies, and private sector firms to develop a next generation Internet for research and education), evolves. This summary is drawn from Internet II project materials.

The technical objectives of Internet II include:

* Maintain a common bearer service to support new and existing applications
* Move from best effort packet delivery to a differentiated communications service
* Provide the capability of tailoring network service characteristics to meet specific applications requirements
* Achieve an advanced communications infrastructure for the Research and Education community

In a number of technical meetings and workshops over the past several years, faculty members and other university representatives have identified a set of advanced applications that will greatly enrich teaching, learning, collaboration and research activities. A major impediment to the realization of these applications is lack of advanced communications services. The broad use of distance learning will require selectable quality of service and efficient "one-to-many" data transport in support of multimedia and shared information processing. To support world class research on a continuing basis, the academic community requires high capacity and selectable quality of service to make effective use of national laboratories, computational facilities and large data repositories.

Internet II is designed to provide a variety of services "on demand" insupport of advanced applications. These dynamically selectable services will include guaranteed bounded delay, low data loss, and high capacity. For example, in order to support delivery of advanced multimedia teaching materials from a digital library repository to a dispersed audience of learners, it will be necessary for the service delivery infrastructure to support "multicast" data delivery with guaranteed upper bounds within the transport components on delay and data loss. New protocols to enable this functionality have already been defined and will be deployed early in the Internet-II project. These protocols include the IETF defined quality of service protocols such as RSVP and RTP along with IPv6, the IETF-developed replacement for the version of IP that is in current use on the Internet. In addition, Internet-II will provide access to the underlying network infrastructure for those environments that can support that access and for those applications that can make use of specific capabilities offered by the infrastructure.

At the heart of the Internet-II design is a new technology for providing advanced communications services. The technology, referred to as a GigaPOP, is a complex of technologies developed over the first decade of the Internet integrated with new technologies developed by vendors and the Internet Engineering community. The Internet-II project will demonstrate proof of concept of this new set of technologies and services so that they can become the basis for the next generation of commercial Internet service offerings. The GigaPOP is the point of interconnection and service delivery between one or more institutional members of the Internet-II development project and one or more service providers. Typical institutional connections will be made via ATM or SONET services at very high bandwidth (at a minimum, OC-12, or 622 megabit/second link speeds and switched data streams as well as packet data routing). The fundamental advance represented by the GigaPOP architecture is dynamically acquired "quality of service" in support of a broad range of new applications while maintaining a common interoperable "bearer service". Service characteristics will include end-user definable capacity as well as latency. An essential part of the Internet-II project will be to determine the incremental costs associated with support of differentiated classes of service and to develop the mechanisms to collect data about the use of

these resources by individual users.

One or more wide area communications service providers will connect to the GigaPOPs in order to provide communications paths between the nationwide set of GigaPOPs and between GigaPOPs and the established commercial Internet. Thus, participating institutions will be able to acquire a wide variety of commercial as well as pre-competitive communications services over a single high capacity communications link to the nearest GigaPOP facility. In particular, to support high performance distance learning and remote collaboration initiatives, the GigaPOP architecture will facilitate local interconnectivity between the higher education community and those commercial providers offering emerging high-bandwidth home access technologies.

Electronic Document Standards

Carl Fleischhauer, Technical Coordinator, National Digital Library Program, and his colleagues at the Library of Congress have prepared a set of papers to assist applicants for the Ameritech awards which together form an excellent basic overview of the situation in standards for electronic documentation. They can be found on the LC site on the web, with the address prefix http://lcweb2.loc.gov/ammem

* Access Aids and Interoperability: (/award/docs/interop.html)
* Digital Historical Collections: Types, Elements, and Construction" (/elements.html)
* Digital Formats for Content Reproductions (/formats.html)
* Identifiers for Digital Resources (/award/docs/identifiers.html)

The papers cover developments in standards, such as the Dublin initiatives, EAD, SGML, PURL's etc. through the late summer of 1996.

CONTRIBUTORS

David Bearman has been President and Senior Consultant at Archives & Museum Informatics in Pittsburgh, Pennsylvania, since 1986. Previously, he served as Deputy Director of the Smithsonian Institution Office of Information Resource Management and Director of the National Information Systems Task Force of the Society of American Archivists. He co-chaired ICHIM '95 held in conjunction with the Museum Computer Network in San Diego in October 1995.

Phil Bantin and **Gerry Bernbom** are Co-Directors of the Indiana University Electronic Records Project. Bantin is Director of theUniversity Archives Program, and Bernbom is Associate Director and SeniorIT Architect, Office of Information Technologies.

Ecaterina Geber is from Eastern Europe. She works at ARTEXPO Foundation in Bucharest, Romania, an independent, non profit, non governmental organization, dedicated to the promotion and development of visual arts, printing crafts and modern communication technologies, supporting, through joint projects in the country and abroad, programs to enhance the impact of visual arts.

