Metadata and the Archival Management of Electronic Records: A Review

by **DAVID A. WALLACE***

Résumé

Plusieurs des plus prolifiques et importants archivistes spécialisés dans le domaine des documents informatiques ont plaidé pour une approche méta-informationnelle de la gestion des dossiers électroniques. Cet essai présente le concept de méta-information et examine à la fois son contexte et son contenu dans la «deuxième génération» des études relatives aux documents électroniques. Ces écrits émettent l'hypothèse qu'une approche des systèmes méta-informationnels aura des conséquences sur l'évaluation archivistique, la conservation, le classement, la description et la référence. Le survol de quatre importants programmes de données électroniques met en relief leurs applications méta-informationnelles et met en évidence le bien-fondé de cette approche pour d'autres archivistes. Une approche des systèmes méta-informationnels va devenir la stratégie fondamentale des archivistes chargés de la gestion des systèmes des documents informatiques. Pour être fructueuse, cette stratégie va requérir le dynamisme des archivistes, qu'ils développent de nouvelles aptitudes technologiques et s'engagent dans la multidisciplinarité.

Abstract

Several of the most influential and prolific electronic records archivists have advocated a metadata systems approach to the management of electronic records. This essay introduces the metadata concept and reviews both its context and its content in the "second generation" of electronic records archives and scholarship. These writings theorize that a metadata systems approach will affect archival appraisal and preservation, and fundamentally alter arrangement, description, and reference. A survey of four leading electronic records programmes examines their metadata applications and underscores the value of this approach for other archivists. It is concluded that a metadata systems approach will become the basic strategy for archivists seeking to manage electronic records systems. To be successful, this strategy will require archivists to become proactive, develop new technological skills, and commit themselves to multidisciplinary collaborations.

Summary

For several decades now archivists have struggled to develop practices and principles for managing electronic records systems. Over the last four years several of the most prominent electronic records archivists have suggested that a metadata approach provides such an operative strategy, and several archival programmes have in fact instituted variations of the metadata theme.

This article examines recent archival writings on metadata and surveys four programmes: the state archives in Kentucky and New York and the National Archives of Canada and the United States, all of which are employing a metadata approach for managing their electronic records systems. These discussions are preceded by two sections. The first is an outline of the problem that electronic records systems pose for archivists, anticipating why a metadata systems strategy is proposed in order to mitigate the loss of institutional memory. The second presents a brief description of data dictionaries as repositories of metadata, examines how other information professionals use data dictionaries and metadata to assist their work, and underscores the value of their perspectives for archivists.

First of all, what exactly *is* "metadata"? "Meta" is a prefix indicating a self-referential term.[1] Hence, a "metatheory" is a theory about other theories. "Metadata," under this simple prescription, is definable as "data about data." However, this definition is both unsatisfying and unclear. In the archival literature, ignoring the standard "data about data" definition, archivists have defined metadata as:

> [d]ata describing data and data systems, that is the structure of databases, their characteristics, location, usage and the like.[2]

Given this definition's presence in three separate archival glossaries, and barring the existence of other formalized definitions, it can be considered the current standard for usage. Recent discussions of metadata in the archival literature offer far more meat to attach to this skeletal outline. Over the past three years electronic records archivists have identified the advantages of a metadata systems approach to electronic records. These include:

(1) capture and preservation of record context (evidence);
(2) preservation of systems and record structure;
(3) generation and retention of relevant descriptive information;
(4) incorporation of appraisal and disposition data;
(5) life cycle management of records;
(6) preservation and migration of system functionality; and
(7) creation of inventory/locator systems for organizational information resources.

These archivists theorize that a metadata systems approach will affect archival appraisal and preservation and fundamentally alter arrangement, description, and reference. Appraisal is seen occurring in the systems design stage, with functionalities built into systems to identify information deemed to be of archival value. Preservation will be accomplished through adoption of the developing Information Resource Dictionary System (IRDS) standard, which will facilitate both migration of entire systems (data plus metadata—ensuring continued linkage between contextual information and record content), and access to resources residing on different software and hardware platforms. Provision for arrangement and description will occur during the systems design stage as well, since post hoc attempts to execute these functions will collapse under the sheer weight of the mass of data contained within the systems and fail to produce adequate finding aids.

The precise impact of a metadata strategy remains unclear, however, given that too few archivists have addressed or debated this issue, much less attempted to apply it to their own operations. What is abundantly clear, nevertheless, is that archivists must become involved in multi-

disciplinary "frontend" systems design projects and must be appropriately educated in order to be able to do so. Such proactivity provides the profession with a strategy to ensure that record integrity meets archival standards.

Up to this point archivists have concentrated solely on capturing "static" metadata that provides information on system content and structure. While this is crucial and necessary, future metadata applications will need to consider the value of capturing "process" metadata in order to provide information on the underlying social and business processes which are written into software. This metadata will explain why a system administers information the way it does in support of the organizational role and mission.

Introduction

Terry Cook has recently suggested that archivists have entered the "second-generation" of electronic records archives, where complex systems, including sophisticated database management software, hypermedia, compound and virtual documents, and advanced telecommunications, are threatening the archivist's ability to maintain adequate documentation of important transactions and decision-making.[3] An electronic records management model developed in a recent United Nations report lists six types of information "objects" produced in this new environment:[4]

(1) Structured Items - vouchers, travel orders, invoices, and purchase orders;
(2) Semi-Structured Items - letters, memoranda, telexes, faxes, e-mail, and reports;
(3) Audio-Visual Items - blueprints, maps, photographs, sound recordings, videos, movies, and tables and figures;
(4) Compound Items - combinations of the above three;
(5) Aggregation of Items - files/folders, records series, and mixed-media aggregations; and
(6) Linked Items - databases, annotations, copies, and hyperdocuments.

These documents can be drawn from electronic records systems that are in a constant state of flux. This dynamic environment allows users to make decisions and take actions based upon a "snapshot" view of these systems as they exist at a particular point in time. The archivist's dilemma, therefore, is how to document the "document" upon which a particular decision is made or action taken.[5]

Failure to manage properly these records and the systems that produce them will render archivists powerless to meet one of their primary tasks: the preservation of institutional memory. The challenge posed by electronic media is not only one of preserving the records themselves, but extends to preserving their various relationships—relationships between each other, their creator(s), and their receiver(s) in the communication process. This necessitates the creation and retention of evidence of authenticity, of "record creation, use, alteration, merging, deletion and transmission."[6] Margaret Hedstrom argues that archivists must strive to maintain the connection between context and content, offering organizations "usable, reliable, authentic, comprehensible, and lasting documentation."[7]

The traditional practice of passively preserving the record at the terminal stages of its life cycle will fall far short of what is needed in the second generation. The record may not survive without early identification and intervention and, equally important, may not contain all of the relevant information that paper-based records systems inherently contain. It is not surprising, therefore, that an advanced institute of government archivists listed the management of electronic records as "probably the most important, and certainly the most complicated, issue currently before the archival profession."[8]

Michael Buckland has offered prescient advice, which, while only tangentially addressed to archivists, should be heeded nevertheless. Buckland observes that a significant change in the technology used to handle "representations of knowledge, facts, and beliefs" may significantly

affect information systems and services. In order to understand this impact it may be valuable not only to study the application of this technology, but also to use it as an opportunity to improve archival theory and enhance archivists' understanding of what we ourselves do and how we do it.[9] Electronic records offer archivists an opportunity not only to face the challenge presented by this rich and complex format, but also to "navel-gaze"—to confront prevailing theory and practice.

Cook points out that archivists, by the very nature of their work, possess the conceptual framework for mitigating the challenges posed by modern electronic records systems. Fundamentally, archival work deals with elucidating the context of records and their relationships. It is therefore contingent upon the profession conspicuously to transfer these traits to the electronic realm in order to ensure the preservation of the "evidential and functional context" of electronic records. For without this, records lose both their meaning and their value.[10] Cook argues that metadata can contribute largely towards establishing a *"de facto* contextual agenda among the creators and contemporary users of these new records that is very compatible with archival presuppositions concerning provenance, interrelationships, context, order and evidential transactions."[11] What is needed is a reassertion of what archivists do and a reorientation of how and when we do it. Cook is not alone. Over the last couple of years several authors have gone to great lengths to articulate the value of metadata for electronic records archivists. It is perhaps an appropriate time, therefore, to review these writings and discuss current metadata applications within a few pioneering archival programmes.

The Nature of the Problem

The traditional practice of passively preserving the record at the terminal stages of its life cycle will fall far short of what is needed in the second generation. The record may not survive without early identification and intervention and, equally importantly, may not contain all of the relevant information that paper-based records systems inherently contain. Margaret Hedstrom, in an article on appraisal, has concisely described the difficulty that the information age poses for archivists:

> Contemporary records are too voluminous, their interrelationships too complex, and the time to appraise them too short, to allow archivists to review all potentially archival records on a case-by-case basis. Unless archivists refine and implement new appraisal techniques to shape the historical record as it is being created, appraisal will become limited to evaluating the remnants of record-keeping systems that someone forgot to erase or destroy.[12]

Indeed, archival institutions are ill-prepared to manage the volume of materials that our society generates.[13] Terry Cook has noted the sheer impossibility of the situation faced by a lone archivist at the National Archives of Canada. This individual has, as one-third of her/his duties, the responsibility to document the government function of national job creation and employment services. This activity takes place in 1,000 offices, is subsumed within fifty programmes, annually produces 3,000,000 case files and 100,000 linear feet of records, and houses twenty-three national and over 100 regional databases, which account for approximately 60,000 transactions per day.[14] One could well assume that the United States National Archives fares no better, given agencies the size of the Department of Defense or Education.

In fact, in 1985, the Committee on the Records of Government noted that the National Archives of the United States employed a mere nineteen archivists to appraise the records produced by over 500 government agencies and only five archivists[15] to appraise the government's electronic records, which by the end of 1984 were being created on nearly 19,000 mainframe computers at 4,000 locations and stored on over 11,000,000 computer tapes and on hundreds of thousands of personal computers. Forecasts from the mid-1970s had estimated that by 1983, seventy-five per cent of federal government information would be handled electronically.[16]

The report also noted the ease with which computer records can be erased without a paper copy ever being produced; the problems associated with version control, where drafts are overwritten, losing policy and programme development documentation; and the retrieval errors that are likely to occur through the absence of file-naming conventions. In the absence of archival policy oversight, individuals are likely to act as "file clerk, records manager, archivist, and indexer" at their own computers and use their word processing software as a "typewriter, filing cabinet, copying machine, and shredder." Archival intervention at the earliest stage of the life cycle—systems development—is one strategy promoted to ensure electronic records preservation and understandability.[17]

Currently, the National Archives and Records Administration does not accession electronic records stored on floppy disks, tape cartridges, or optical disks. Instead, it requires that agencies transfer electronic records onto magnetic tape, and provide appropriate documentation. Transferred files must be converted into a hardware- and software-independent format, in either Electronic Binary Coded Decimal Interchange Code (EBCDIC) or ASCII.[18] To date, NARA's Center for Electronic Records has received less than ten per cent of the records that agencies originally agreed to send.[19] A 1991 report commissioned by NARA identified hundreds of previously unregistered databases, and also found that agency database management too frequently did not adequately consider archival concerns. The investigating body often found systems documentation "skimpy or nonexistent."[20] Noting that agencies are developing more complex integrated systems, it suspected that many new systems are neither inventoried nor covered in NARA record disposition schedules. Furthermore, it was concerned that agency systems design continues to ignore preservation and archival requirements:

> [t]he findings of the project staff and the discussion among expert group and panel members point strongly to the need for major emphasis on this integration to assure preservation of important electronic records in the future, and the long-range solution to the problems which exist today.[21]

A box of textual records can lie in a corner for twenty years and still be evaluated and, if judged to be of archival value, accessioned. Electronic records provide no such luxury. Even if by some odd chance the media are readable (which is highly unlikely due to magnetic instability and hardware obsolescence), without appropriate documentation the data are uninterpretable. Loss of software functionality means loss of contextual information regarding how the data were represented and manipulated.

How well prepared are archivists to undertake the challenge posed by electronic records? In his doctoral dissertation, Richard Cox summarized the record of archival research and publication in the following manner:

> The cumulative results of an archival literature that is over a half-century old is that the American archival profession still lacks a basic knowledge about the origin, nature, reasons for success and failure, and other characteristics of archival repositories.[22]

The relatively underdeveloped nature of archival education, especially in the United States, is another structural limitation inherent in the profession which has inhibited the implementation of uniquely archival strategies for managing electronic records. In a recent survey of graduate archival education programmes in North America, Cox found that out of 227 course offerings, only six specialized courses existed that dealt specifically with electronic media (four on automation techniques and two on electronic records).[23] Nearly two decades ago, when some archivists were envisioning the potential impact that computer technology would have on society, a United States National Archives course offering on electronic records was cancelled due to lack of interest.[24] The result of this negligence has been that archivists are poorly trained and politically too weak to intervene substantively in the design and use of information technology within organizations. Consequently, "individuals and organizations have developed their own

conventions for handling information in automated systems [and] many current institutional practices undermine retention, preservation, and secondary use of electronic records [because they are often] inefficient, non-standard, and difficult to reverse."[25]

One of the reasons commonly offered to explain archivists' poor response to the challenge of electronic records management has been that traditional practice for managing more conventional and familiar formats (such as paper and audiovisual materials) has nearly completely focused attention on the termination of the life cycle. Consequently, archivists have invested most of their conceptual labour on this stage. What is abundantly clear from the above discussion is that archivists cannot treat electronic records in the same manner, and that new strategies are called for. Experienced electronic records archivists are adamant in their determination to reorient archival intervention to the earliest stage of the life cycle: systems development.[26] In addition, a modified life cycle has been developed to assist in the management of electronic records:

(1) Creation and Identification - capture and store records, assign unique identifying information, provide security, and support authentication clearance, and version control;
(2) Appraisal - maintain and apply appraisal criteria and authorization;
(3) Control and Use - facilitate internal/external record communication, search and retrieval, provision of tracking mechanism, and facilitate migration, conversion, and portability; and
(4) Disposition - notify creator/user of disposal, dispose of records identified by retention schedule, maintain archival information in independent format.[27]

Archivists who have discussed metadata understand that what is needed is a proactive outlook. For it is through the provision of metadata capture and creation that archival functional requirements[28] can be satisfied. They provide an operative strategy to face the dilemma posed by the mass of information and the survival of records (and their associated contextual information) and records systems processes. Data dictionaries are one tool available to assist archivists in this regard.

Data Dictionaries and Metadata

In the electronic records systems environment it is appropriate to distinguish between the structure imposed by the software and the information entered by the user. The information entered by the user is the data, whereas information on the rules for structuring and presenting this information, as defined by both the system and the systems personnel within the software, is the metadata. In this environment there are two types of metadata: one describing what the data is and what it means; and the other indicating where the data can be found and how it can be retrieved. The first is data dictionary metadata; the second is directory systems metadata; the latter is subsumed within the former.[29] A metadatabase is a collection of metadata managed and controlled as a discrete unit. It provides a facility to organize, access, and control metadata.[30]

Metadata housed within a data dictionary is built up as the database structure/architecture is developed and altered. Simply defined, a data dictionary is the "central storehouse of data about an information system's data and the data's transformations,"[31] and describes the "logical aspect" of the data: how it is presented and seen by the user.[32] More specifically and concretely, Rom Narayan views the data dictionary as both a "tool and a resource." As a tool it enables organizations to

> document, organize and control an organization's information resources. (Including manual and automated data, such as data elements, reports, screens, forms, etc., and all methods used for conveying information, such as programmes, processes, machines, networks, transactions, etc.) It is a tool for people who need to know what the information processing resources of the organization are (typically, data processing resources such as pro-

grammes, databases, and application systems) and how these resources are used by different users. As a resource, the dictionary is an organized repository of information describing the source, use, edit criteria, control, user responsibility, and content of data within an organization.[33]

The dictionary is composed of two parts: the database and the application system. The metadata is stored within the database and the user interface is contained within the application system. The data dictionary can be used:

(1) to document data, users, procedures, and how they interact with one another;
(2) as a system development tool, [to] provide
    - Copy files for consistent presentation of data, and
    - Integrat[ion] in systems development;
(3) as a data administration tool for the control and integrity of the data;
(4) as a planning tool to predict the impact of changes; and
(5) as an inventory of resources to identify where information is located.[34]

Narayan lists several advantages of the data dictionary. These include:

(1) Organization of data
    - Sources and uses
    - Relationships
    - Physical characteristics
    - Identification of users
    - Definitions of "where kept"
(2) Control of data
    - Naming conventions
    - Standardized processing
    - Operations scheduling
    - Data security
(3) Integrity of data
    - Standardized definitions
    - Assigned responsibility for data items
    - standardized validation rules
(4) Basis for inquiry and analysis[35]

What is interesting and sobering is that archivists are absent from the group of users that Narayan sees benefiting from a data dictionary system. In lieu of the archival profession, systems developers, systems analysts, application programmers, database administrators, operations staff, managers, and even end-users are listed. This is perhaps not surprising, since most of the literature on data dictionaries appears to be oriented towards business and corporate computing. However, even a cursory reading of the extensive list provided by Narayan should make it apparent to archivists who work with electronic records that their own work can be enhanced immensely through data dictionaries and the metadata they contain. It is metadata that renders the data understandable and locatable, serving simultaneously as a description and locator device. Lacking it, the data exist without structure and remain uninterpretable and unretrievable. Metadata is intricately interwoven with the data and an organization's ability to manage its information resources. It provides the means for representing the data as it was originally seen by its creators and users. Without the metadata, the data lose their representational *form*, which, according to Peter Sigmond, may offer an understanding not only of organizational function but also of a document's value for archival preservation.[36] Likewise, Bruno Delmas has noted that elements of document form can be captured in metadata and contribute to an understanding of organizational function.[37]

Contemporary automated data dictionaries are capable of collecting and creating metadata documentation on:

(1) Data fields - including information on name, alternate name, type and length, output format, default value, field caption, source, access privileges, validation rules, calculation rules, and text description;
(2) Data flows - pathways by which data travels from one location in an information system to another. Contains information on name, alternate name, abbreviation, text description, origin, destination, and volume and frequency of flow;
(3) Data stores - repositories for data that remains idle for some length of time during a data flow process. Contains information on name, alternate name, abbreviation, record name, text description, input data flows, output data flows, and volume and frequency of flow;
(4) Processes - alters or changes data from one form to another. Contains name, purpose, input data flows, output data flows, process description;
(5) External entities - person, department, external organization or information system that either provides data to, or receives it from the system. Contains name, alternate name, acronym, input data flows, output data flows, and description;
(6) Records - meaningful combinations of data fields. Contains name, alternate name, text definition, and data element content (those data elements included in the record); and
(7) Reports - alphabetical listing of all data elements, listings of all data flows and data stores that include a particular data element, or other more specific listings.[38]

Any archivist can see that if the data dictionary can contain these types of useful metadata, it can also be designed to hold documentation relevant to the archival endeavour, such as appraisal analyses (archival value, retention and disposition schedules), provenance data (name, function, mission), audit trails, distribution lists, version controls, and access restrictions.

Data dictionaries can also be used, with the assistance of computer-aided software engineering (CASE) tools, to construct data flow diagrams (DFDs)—graphical representations that provide information on how data travels through an information system. These models can be used to track "processes" through which data can move from one source to another, be operated on, possibly transformed, and returned or moved on to a third source.[39] This is similar to the archival concern for tracing a document's movement within a records system, movement which bears directly on a record's value as evidence of organizational process and/or function.

For example,[40] let us imagine a cable coming from an overseas embassy into the Department of State, where it is read by the Secretary of State, who annotates it and attaches a information memorandum to it, then passes it on to the President for review and possible action. The original cable is housed with the originating embassy; a copy is passed to the Secretary of State, who annotates it and attaches the information memorandum. This new aggregate original is placed in his/her files and a copy of it sent to the President. The President in turn attaches an action memorandum to it, files this new compilation within the White House's record-keeping system, and sends the expanding aggregate document to the National Security Council for further action. The National Security Council receives a copy of the annotated cable, the Secretary of State's information memorandum, and the President's action memorandum. One can easily see from this small example the value of process-tracking offered by data dictionaries. While the development of such a graphic representation is necessarily complex, it encourages archivists to concentrate on the processes that produce the record rather than focusing on the records alone. JoAnne Yates encouraged archivists to do precisely this in her examination of communication systems in United States businesses between the years 1880 and 1920.[41]

Discussions on the data warehouse, an accumulation of integrated, subject-oriented databases created to assist managerial decisions, explicitly address the benefits of metadata in a manner familiar to archivists. In this environment metadata is used to control externally-generated, unstructured data within the warehouse. Typical metadata collected includes document ID,

entry date into the warehouse, document description, document source, document classification, indexing, purge date, physical location, document length, and related references. It is further noted that "archival" data in a data warehouse must be stored with its metadata, since "nothing is more frustrating than...trying to solve a problem using archival data when [one] does not know the meaning of the contents of a field being analyzed."[42]

Data administrators have also commented upon the value of data dictionaries and metadata to their work. Data administration "assists [in] the provision of information systems by controlling and/or coordinating the definitions (format and characteristics) and usage of reliable and relevant data."[43] Data administrators are given responsibility and control over data dictionaries, which are viewed as their primary tools since they store "all the information about data, databases, programmes, reports, panels, business functions, and other project details" that enable an organization's information to be treated as a valuable resource.[44] Data administrators' concern for centralized control over data definitions and their usage is compatible with archivists' growing concern over maintaining adequate centralized documentation of electronic records systems in locator systems. This "front end" processing is understood by data administrators as a means of reducing later efforts, such as "documentation and auditing of systems,"[45] which are exact concerns of archivists responsible for any record format.

Archivists have much to learn from other information professionals in managing electronic records and electronic records systems. Data dictionaries and the types of metadata that they house *and can be built to house* should be seriously evaluated by archivists, for their potential value to the profession lies beyond those issues outlined above.

Archivists and Metadata: A Review of the Literature

Metadata, as it has been defined and employed by archivists, incorporates many traditional archival functions. In this sense, archivists do not merely acquire metadata in order to document electronic records systems in the ways described in the discussion of data dictionaries; they also need to create metadata in order to provide contextual information about the data.

David Bearman has aptly noted that archival descriptive systems have always been metadata systems: "systems of information describing information systems."[46] Through this lens the term "metadata" becomes less abstract, and the concept it embodies is recognizable to archivists.

Active acquisition and retention of information that traces records use ensures preservation of the data's evidential context, which in turn ensures "meaningful access."[47] In paper-based systems this functionality is often preserved naturally and is evident in the physical record itself.[48] However, in electronic records systems, this information must be preserved through the active intervention of the archivist and acquired as part of a metadata documentation strategy. Richard Kesner sees metadata as capturing the crucial evidential value of records, whereas the data itself merely represents its informational value.[49] This point is well taken, for without its contextual information a record's value as evidence of provenance, function, or action remains unmet. Furthermore, the accumulation of an organization's information resources in a metadatabase provides the means through which information systems can be identified, inventoried, and located.

The ongoing strategic agenda of the National Historical Publications and Records Commission (NHPRC) for electronic records management has recognized the tactical importance and value of metadata for more than three years now. In 1990 it made as its first recommendation the requirement that archivists include "archival components" within electronic records systems, in order to "assure the preservation of historically valuable information." Of the means proposed to accomplish this task the report suggested a specially-designed archival application of the Information Resource Directory System (IRDS) to store metadata.[50]

The following year, NHPRC issued a report laying out a specific research agenda for electronic records. No longer subsumed under a broader recommendation, metadata systems were acknowledged in and of themselves as an integral research area. Proposing a multidisciplinary approach, the report asked archivists to consider how such systems could "be used to support electronic records management and archival requirements." Recognizing that existing metadata systems such as data dictionaries, Information Resource Directory Systems (IRDSs), and inventory/locator systems, already house some of the types of information that archivists need to describe and control electronic records, the report encouraged archivists to explore how these systems can be further exploited to include new types of information with which to support archival functions and requirements.[51]

Finally, recent guidelines for NHPRC electronic records research grants not only reiterate the ten research questions identified in the 1991 research agenda, but also point out that its support of research undertaken by the New York State Archives and Records Administration (SARA) has funded a pilot project to test the transfer and preservation of metadata about electronic records.[52]

The Report of the SAA Working Group on Standards for Archival Description (WGSAD) has noted that modern electronic records systems contain much "self-referential"[53] information that assists the descriptive process and encourages archivists to exploit this data in lieu of recreating it as records are appraised for archival retention. The report's redefinition of "archival description" reflects this new outlook and the place of archivists in the life cycle of information:

> Archival description is the process of capturing, collating, analyzing, and organizing any information that serves to identify, manage, locate, and interpret the holdings of archival institutions and explain the contexts and records systems from which those holdings were selected.[54]

The relevant archival literature on metadata and metadata systems is found in writings published over the last three years. One of the earliest and most comprehensive analyses is found in Chapter II of the United Nations report entitled *Management of Electronic Records: Issues and Guidelines*,[55] which promotes metadata systems as a "primary tactic for control of information about records."[56] Arguing that, given the volume of records produced in the computing environment, records managers and archivists must develop a strategy for managing these records in an aggregate manner, the report challenges archivists to focus upon "application systems"—those information systems that assist communication and work—and "systems management," the means to oversee the life cycle of records resulting from the application. The report contends that records managers and archivists must devise methods to "sculpt" application software and records transactions. They must do so in order that the resulting records accurately reflect organizational roles and meet the demands of organizational accountability. Consequently, records managers and archivists must intervene in systems design in order to ensure that documentation is retained which demonstrates not only what was communicated, but also how the information was used.[57]

In order to identify and document the information systems within an organization, the report advocates the use of Information Resource Dictionary Systems (IRDSs). These metadatabases, normally developed by systems personnel as part of their general duties, can be enhanced by records managers and archivists through inclusion of long-term retention documentation.[58] The report argues, however, that archivists still need to examine how metadata systems can be fully exploited within existing archival descriptive practices, a point reiterated in later writings.

Configuration management—the maintenance of a record tracing system changes—is offered as a means for preserving the changing contexts within which records are created and used. It provides a means for recalling earlier database structures in order to examine the various views offered to users at particular points in time.[59]

David Bearman, who authored the above-cited sections of the United Nations report, has forcefully argued that archivists need to recognize that records foremost are evidence of transactions that document organizational functions and provide accountability. Without this conceptual framework, archivists will be lost in the electronic systems environment, believing instead that what they need to accomplish is identification of "types" of records to retain rather than documenting important transactions. Not all relevant transactions lead to records creation in the traditional sense—i.e., result in an identifiable and recognizable entity. For example,

> searching a database in order to generate reports may be an important decision making process but it doesn't generally lead to creation of an electronic record or even assure the preservation of the particular "view" of the data or the analytical or reporting models being employed in the presentation.

In this environment the data alone are inadequate for preserving evidence of transactions. The requirements for evidence are met through the integration of system structure, context, and data. Given this situation, Bearman argues that it is imperative that archivists be involved in systems design in order that these data management requirements are provided for, since content alone tells us little about how a record was used or how it was seen at the time of creation. Metadata systems are considered as assisting archivists in preserving record context and system functionality through retention of information such as database views, analysis and reporting capabilities, security provisions, and processing routines. Metadata systems will also aid reference and access by serving as finding aids to the systems resources. In order for archivists to accomplish their mission in this area, however, Bearman argues that they must co-opt information technology experts onto their staffs, in true interdisciplinary projects, rather than presuming to learn all the skills maintained by these professionals.[60]

A follow-up United Nations study has constructed an Electronic Records Management (ERM) Information Model, designed to deal with information objects created by modern information systems. In order for objects to be processed (identifying "*who* performs *what* functions on *which* information and *when*"), they must be classifiable. The ERM Information Model identifies two types of attributes which assist classification: content attributes and metadata. These have been defined as follows:

(1) Content Attributes
- access conditions/restrictions permissions
- authentication/concurrence/clearances
- date
- record item type
- identifiers
- incoming/outgoing
- language
- substantive/facilitative
- to/from
- version number

(2) Metadata
- active/inactive (current/non-current)
- appraisal decision
- arrangement (collections only)
- structure of components for compound documents
- format (logical relations)
- frequency (recurrence)
- history of use (audit trail)
- links and pointers (subject terms, hyperdocuments)
- ownership
- structure (physical layout/presentation)

Several of the content attributes listed above are recognized as metadata attributes in other archival writings. Since this model offers prescriptive advice for software functionality, which approximates archival concerns for context evident in all archival writings on metadata, it should be evaluated by archivists and its application by the United Nations closely followed.

Management of information objects is one of four activities that are considered basic for ERM by the United Nations The other three are related to internal architecture, transfer, and storage.[61] An Information Resource Dictionary System (housing metadata) is seen as contributing to the Management of Information Objects segment of ERM.[62]

In his doctoral dissertation, Hamza Kandur recognizes the value of metadata for capturing information confirming record authenticity. He argues that an IRDS can provide information on a user's actions relating to records creation and alteration. Such a capability is seen as essential for organizational accountability. Archivists can contribute substantively to organizational functioning through creation and retention of this type of integrity-related metadata. Appraisal metadata captured at the system design stage within an IRDS can mitigate labour-intensive collection of this information at a later stage.[63] Indeed, provisions for comprehensive documentation and description of records and information systems at the design stage will provide the basis for intellectual control, and minimize requirements for later redescription.

Kandur also comments on the value of metadata for locator systems. Building on the model presented by the National Archives of Canada of a suitable data dictionary for managing electronic records, Kandur lists the following metadata components of an ideal directory for accessioned data files[64]:

* Creator - person, organization, or organizational unit
* Title - of the file
* Dates - of creation and modification
* Classification number - a unique identifier
* Series information
* Addressee
* Data collection - information on how the data was collected, the methods of collection, time span, etc.
* Content - summary of file contents
* Related sources - references to related records
* Accompanying material - i.e., input forms, questionnaires, etc.
* Restrictions on use
* Type of data - textual, numeric, survey, etc.
* Size - of file
* Technical requirements - storage medium, hardware, software, operating system
* Retention information
* Disposition information

Kandur concludes by contending that the problems posed by electronic records are related more to management than to technology. Archivists and records managers must communicate their requirements better and create new working relationships with information technology managers.[65]

Charles Dollar's volume on information technology and archival theory asserts that the principle of provenance will continue to be the cornerstone for intellectual control and preservation of electronic records.[66] Dollar notes that in traditional records systems, the physical document by its very nature contains "explicit context attributes" such as date, originator, and addressee. Therefore, the document's physicality is itself evidence of a transaction. Electronic records, however, do not necessarily contain these attributes, since the record is a logical rather than a physical entity. It only becomes physical once it is printed or is displayed on a terminal through software control. Unfortunately, not all software guarantees that sufficient "context attributes"

will be captured (as evidenced by the United Nation's explicit provision for these kinds of data elements in its information model). Given these complications, Dollar argues, as do Bearman and Kandur, that it will be metadata systems such as data directories that will be responsible for capturing much of the context information (such as provenance) that archivists will need in order to interpret electronic records systems. Consequently, Dollar is adamant that archivists must become active players in the development of the Information Resource Dictionary System (IRDS) standard in order to ensure that archival requirements are understood and adopted within it.[67]

The IRDS standard is currently a draft international standard that will allow an organization to build a single seamless resource directory regardless of the software and hardware platforms on which an institution's information resources reside. As currently configured, the IRDS would include a

> definition of the data an enterprise requires, a description of the processes available for delivering and maintaining the data, a delineation of both the hardware/software environment in which these processes function and how the data is intended to be used, and an identification of the individuals and sub-organizational units and function associated with specific applications or processes. [For archivists and records managers the IRDS could be crafted to address] provenance, evidential value, audit trails, disposition, access, privacy, and the like.[68]

Dollar asserts that archivists should also build provisions for the life cycle management of electronic records directly into metadata systems, so that information suitable for the identification, retention, and retrieval of records of continuing value may be captured. The advantage offered by the IRDS is that it will enable whole systems, including the data and their defining metadata, to be migrated onto new systems platforms. It is argued therefore that archivists will need to concentrate their efforts on metadata systems creation, rather than informational content descriptions, since in the electronic realm, archivists' concern for informational value will be eclipsed by concern for the evidential value of systems. Dollar implies that this is one of the fundamental shifts wrought by electronic information systems on the way in which archivists will operate.[69]

Other basic archival functions will be altered as well. Arrangement and description will be collapsed into a single activity, largely performed during systems design, before the records are even created. As noted in the previous paragraph, description will focus on the information systems context, in support of information-sharing across the organization, and not on the content of the records themselves. The resultant metadata product will serve as both an inventory and a finding aid. Additional value could be added later by both archivists and users of the system. The projected metadata system (ideally within the IRDS standard) would identify

> all of the information elements, define their relations, explain their context of creation and use, provide audit trails of use, and specify organizational responsibility for their maintenance.[70]

Reference and preservation would also be affected by a metadata systems approach. Reference will occur through the IRDS, which will house information on all of the information resources—even those housed on incompatible software—within an organization, and will act as a locator for this information. Preservation will be enhanced by the IRDS, since it will allow entire systems (data plus metadata) to be migrated onto new systems, ensuring intelligibility by retaining linkages between contextual information and the actual contents of the records.[71]

Dollar closes by stressing the point made earlier by Cox, namely that the electronic records systems environment requires archivists to acquire new skills, which existing archival education programmes have neglected imparting. Dollar would focus this education on operations research and information systems design.[72]

In a paper given at the 1992 Annual Conference of the Association of Canadian Archivists, Margaret Hedstrom encouraged archivists to envision a new methodology for archival description of electronic records, a methodology that respects archival principles but at the same time takes advantage of the new technologies.[73] Drawing on some of the same concerns expressed by Dollar, Hedstrom stipulates that the descriptive standards for electronic records must include provisions for establishing and ensuring record identification, access, understandability, interpretation, authenticity, and ongoing management.[74]

Hedstrom also argues that archivists must learn to take advantage of those metadata systems that organizations already create and incorporate in them archival requirements for inventory and locator systems that capture information on "provenance and context of records creation, [in order to] achieve highly refined access to the contents of records." She asserts that electronic records present archivists with an excellent opportunity to capture organization-created metadata in lieu of a separate archival descriptive process (which she describes as "folly"). By emphasizing systems preservation over data preservation, archivists can ensure the retention of the "organic relation between the content of the records and the ways that organizations structure and describe them." Such a strategy will present archivists with an embarrassment of riches, an abundance of metadata that provides far more descriptive information than the archivist can ever hope to reproduce through traditional methods of description. However, archivists may come to find that metadata management (i.e., the identification of truly valuable descriptive data) is far more complicated than managing archival records alone.[75] By positing that archivists' attention will "shift from creating descriptive information to capturing metadata and managing it to promote access, use, and understanding of archival records," Hedstrom envisions a metadata systems approach (much like Dollar, Bearman, and Kandur) which fundamentally alters traditional archival description and reference. She is quick to point out, however, that metadata is not a panacea ameliorating the difficulties of the electronic age. Archivists have themselves yet to define what types of metadata are essential to meet archival descriptive requirements. These, however, are more cautionary flags than serious reservations. It is apparent that Hedstrom, like the others, sees the metadata systems approach as offering a rich opportunity to develop associations between record content and the context within which records were created and used.[76]

Finally, a recently published volume of essays from a symposium held in Marburg, Germany, in October 1991 has added to the archival discussion of metadata.[77] The editor, Angelika Menne-Haritz, asserts in her introduction to the book that an understanding of records context is "indispensable" for a true understanding of the content of the records themselves, and that it is archivists who more and more are bearing the responsibility for preserving the "comprehensibility of the contextual circumstances" surrounding records creation. Adequate documentation is essential for capturing the contextual and transactional information that raises the document to the level of evidence. It is held that metadata systems designed to capture this type of structured information can help archivists fulfil their role.[78]

Richard Barry, touching upon points made earlier by Cox, Dollar, and Bearman, argues that archivists and records managers must possess a broad range of human, financial, and technological skills and resources in order to take a leading role in electronic records management. They must be educated in order to understand the structure of sophisticated computing systems if they hope to integrate archivally relevant metadata into systems design. They also need to build links with systems-design staff in order to participate in the structuring of metadata systems that contain adequate information on scheduling, appraisal, and disposition. If archivists falter, the vacuum will be filled by others who may be less concerned with preserving records of archival value.[79]

In the book's closing essay, David Bearman underscores the importance of the relationship between provenance—the special concern of the archivist—and a record's value as evidence. Bearman is both fearful and hopeful. He argues that, while provenance information can be

more easily and fully captured in electronic records systems, ironically, the likelihood of the loss of this data may perhaps be greater. A loss of provenance data will render records useless for evidential purposes. As he has argued elsewhere, Bearman here stresses that the combination of a record's structure, context, and content is crucial for an understanding of its creation and use ("evidential historicity") within the organization. In traditional paper-based systems the existing filing scheme (chronological or subject, for example), record-keeping practices, and the physicality of the records themselves (handwritten annotations, ink types, record form) provide us with the valuable contextual information for interpretation. However, electronic systems do not automatically retain or provide for the capture of these types of information.[80] If archivists so choose, they can incorporate, through systems design, provisions for "track[ing] every modification made to a record during its drafting life, every recipient of an electronic message and when they read it, and even every time a document was consulted." However, as similarly noted above by Hedstrom, problems arise regarding precisely what "amount of evidential historicity is required and how to capture and retain it," as well as how to overcome the difficulties associated with hardware and software dependency and migration. Bearman suggests that archivists can and should explore the traits inherent in different software applications (such as word processing, electronic mail, spreadsheets, database management systems, and graphics software), which provide clues to evidential historicity. Where these are lacking, the archivist should strive to construct them in order to fill the gap. The contents of records alone are of no use; what is needed is documentation of structural and contextual metadata from the system within which the data reside.[81]

Overall then, electronic records archivists have discussed the value of metadata systems in some detail. The dominant attributes and advantages of the metadata systems approach articulated by this handful of archivists include:

(1) capture and preservation of record context (evidence)
(2) preservation of systems and record structure
(3) generation and retention of relevant descriptive information
(4) incorporation of appraisal and disposition data
(5) life cycle management of records
(6) preservation and migration of system functionality
(7) creation of inventory/locator systems for organizational information resources

Archivists have yet to determine specifically which types of metadata to create and capture, and how different software applications alter the types of archival functional requirements to be handled by metadata systems—as well as identify which can be accommodated through policy. It has been suggested that traditional archival functions such as appraisal, arrangement, description, reference, and access will be altered by a metadata systems approach; the precise impact, however, remains unclear and too few archivists have either addressed or debated this issue. Nevertheless, it is undeniable that, in order for archivists to manage electronic records effectively, they must become involved in multi-disciplinary "front-end" systems design projects and must be appropriately educated to do so. Such proactivity provides the profession with a strategy to ensure that record integrity meets archival standards.

Nearly ten years ago, Richard Kesner admonished archivists for not being aggressive enough. He argued that archivists should emphasize serving their parent institutions, rather than vague and unknown future users, and free themselves from their traditional passivity, rather than waiting for the records to arrive at the archives. He encouraged archivists to become "information specialists" who take an active role "in the creation, distribution, and preservation of information."[82] Such a notion is not so new or so heretical as many archivists would presume. As early as 1956, Margaret Cross Norton suggested the same approach for government archivists. Given archivists' understanding of administrative history and the potential value of records beyond their operational utility, Norton posited that archivists should take an active role in the creation of records—since it is archivists who have "some definite ideas as to the desirable

qualities for records, both as to their factual content and their physical format."[83] Metadata strategies for electronic records, while reorienting the archivist's role, provide the profession with a tool for accomplishing one of its most important missions: the provision of *understandable and useful* documentation (understandable through the retention of the crucial contextual information that makes their interpretation possible, and useful through the early identification and preservation of records of archival value)—thus ensuring the retention of institutional memory.

Metadata Applications

At least four archival programmes are actively acquiring and creating metadata. They are the Kentucky Department for Libraries and Archives, the New York State Archives and Records Administration, the National Archives of Canada, and the United States National Archives and Records Administration.

Kentucky Department of Libraries and Archives[84]

In 1983, the Kentucky State Archives began exploring a mechanism to define data elements within agency databases, for the purpose of compiling a data dictionary containing an inventory of databases in the state government and information on electronic files. The purpose of this project was to enable the archives to get a sense of the quantity of electronic information held by state agencies, and to develop a strategy to target systems containing records of archival value. The archives was then (and remains today) more concerned with inventorying these systems than with accessioning them within the archives—an impractical task given the available resources.

Until 1982, individual state agencies planned and developed their own computer systems, resulting in minimal uniformity and coordination. An attempt at this time by the Department of Information Services (DIS) to centralize systems development was hampered by inadequate staffing and funding. Given the DIS's mission, little attention was paid to identifying historically important databases for preservation.

Since 1984, the project has been integrated into a larger data management project involving other state information agencies, including the Archives. Known as the Kentucky Information Systems Commission (KISC), this body was charged with overseeing agency automation. Within this framework, the Archives attempted to adapt existing records management policy to electronic records systems.

The initial data dictionary project failed to achieve the type of online inventory of agency databases that it originally sought to capture, due to inadequate staffing and skewed implementation; the data dictionary was operating before a planning group had identified those functions that it should have been designed to support. As a consequence, the Archives switched tracks and sought to capture life cycle tracking of files and information on the state's information resources through the use of three tools: the data dictionary, a public records management database, and a locator metadatabase of information on systems and manual files.

In 1986, an information policy called the Kentucky Information Systems Architecture provided KISC with the ability to review agency data processing. Within this framework, agencies were required to prepare and submit biennial data processing plans, including information on statutorily-mandated agency missions and functions, and expenditure forecasts for existing and planned automation projects. Project statements included commentaries on the envisioned impact of automation plans on agency record-keeping. This data gave the Archives a better picture of existing and planned agency systems than it had previously possessed, and led to the scheduling of several electronic records systems. Armed with these new sources of information, the Archives Public Records Management System database described hundreds of sys-

tems and linked the descriptions with information on existing retention schedules for records in a non-electronic format. This provided the Archives with a more informed base upon which to make appraisal decisions.

Currently, agencies provide KISC with electronic copies of their automation projects in a structured format that provides for discrete data elements to be parsed from these plans; it is believed that in the near future agency mission statements and strategic data processing plans will be collected in electronic form as well. All of this information will then be available for downloading into the metadatabase.

It is hoped that the online locator metadatabase will someday provide access to systems and manual files such as the data dictionary, as well as to information collected during the strategic planning process, and during series identification, description, and appraisal. The construction of these types of metadatabases is viewed as a key element of public records management. The archivists understand that in order to accomplish this task they must operate within a broader environment, including other information professionals such as data administrators, since an autonomous archival initiative along these lines would most likely fail. The archives is understaffed and sees its mission with the metadatabase as largely one of consciousness-raising in order that agencies may see the value of their electronic systems resources. Chuck Robb at the State Archives has deemed the metadatabase project a qualified success. Data dictionaries have been found to be more technically- than retrieval-oriented, and this problem has been compounded by the fact that information captured and reported by the individual agencies for the metadatabase is not fully standardized.

New York State Archives and Records Administration[85]

The Sourcebook Pilot Project is an NHPRC-funded attempt by New York State's Forum for Information Resource Management to improve access to information resources in New York state agencies. The project tested the applicability of a metadata systems approach as a means of increasing data sharing among separate state agencies. The result is an updatable database designed to provide centralized control over the state's information resources. This "sourcebook" contains information on automated databases and manual records systems in eight state agencies. Long-term plans include the creation of a government-wide information locator system (ILS) and provision of print products to complement the electronic metadatabase.

The pilot project was conducted in eight state agencies and resulted in 148 descriptions, including 118 electronic databases (sixty-nine on mainframes, twelve on minicomputers, thirty-three on microcomputers, and four in other formats) and thirty series descriptions of archival records. Information collected on each source included its name, size, temporal range, purpose, contents, strengths, weaknesses, data collection methods, geographic coverage, fees, access restrictions, and identification of an expert contact person within the agency.

This project has had a significant impact on agency views of both their own information resources and the resources of other agencies. Half of those who participated felt that they had gained a broader understanding of their data and underlying programmes; one-fourth reported that the project had led to greater concern over data quality and staff information handling practices; and one-third gained new respect for the advantages of inter-agency information exchange.

It is believed that the ILS, now in the planning stage, will encourage improved information systems planning, development, and investment; enhance inter-agency data sharing; improve public access; and "support and consolidate administrative functions." Coordination and integration of existing information inventories from the State Library, the Archives (automated systems, records inventories, and schedules), the Forum, and the agencies themselves would serve as the basis for developing this tool.

Margaret Hedstrom, who served as Chair of the Sourcebook Advisory Board, sees in metadata a meaningful concept to apply to thinking about functions, context, systems applications, and records. What archivists can offer to a project such as this is an understanding of data context. Metadata provides a methodology for understanding the relationship between content and context, a concept that is frequently lost on other information and computer personnel. The State is starting to focus on the capture of records (metadata plus data) and not just their data, but has yet to capture this metadata completely in electronic form.

Currently, the Center for Electronic Records in the New York State Archives and Records Administration is surveying agencies for information about the policies, practices, and tools that they follow or have created to administer their own data. The survey is also trying to determine whether or not agencies have data dictionaries for their electronic records database systems.

National Archives of Canada[86]

The National Archives of Canada is developing a framework identifying the functional requirements that are considered necessary in order to maintain electronic records office systems. It is currently the responsibility of the agencies themselves to decide what metadata they need to capture. The Information Management Standards and Practices Division (IMSP) at the Archives provides "front-end" records management advice within the context of the mission of the National Archives to facilitate the management of records and, specifically, to help institutions preserve their corporate memory. It strives to develop standards for practice so that agencies can render their records available, usable, and understandable throughout their life cycle (encompassing planning, collection, creation, organization, retrieval, use, accessibility, transmission, storage, protection, and disposition). These concerns have a direct impact on the "quality" of electronic records in terms of their "integrity, currency, and relevancy." Metadata capture is part of a strategy designed to preserve electronic information in a way that meets these criteria. The strategic priorities of the IMSP are concentrated on those information life cycle activities which:

(1) Are vital to the preservation of the corporate memory (identification, description, classification, organization, storage, and protection)
(2) Directly determine what constitutes the corporate memory (planning, creation, collection, generation, and disposition)
(3) Indirectly affect the preservation of the corporate memory (access, retrieval, transmission, and use)

Through this programme the Archives is attempting to develop a new profile within the government, where it is viewed as an administrative arm of the corporate entity instead of a strictly historical/cultural agency. Echoing Richard Kesner, John McDonald, Director of the IMSP Division, contends that archivists must fulfil the needs of the corporate body before they address future records use by researchers. In this capacity, the IMSP views itself as an integral component ensuring the preservation of the government's "corporate memory," and concerns itself more with what agencies "do to" records to render them available, usable, and understandable, than what they "do with" them.

Phase I of the IMSP's Information Management and Office Systems Advancement (IMOSA) project sought to satisfy the above criteria through development of prototype software to support the management of office documents in both electronic and other formats, and meet functional requirements for managing information within office information systems (including records collection, filing, access control, retention and disposal, migration, formatting, classification, search, and retrieval). The larger goals are to identify and address the functional requirements for managing information as new information technologies enter the office. Phase I concluded that the most important issues for managing such records are not technical, but are

related to human and organizational factors such as policies (accountability or lack of accountability), procedures, practices, perceptions, and behaviours. It also concluded that any information management tool, such as the prototype software, in order to gain user acceptance must be:

(1) comprehensive - covering all phases of the life cycle, regardless of technology or medium;
(2) flexible - able to handle different formats and media and provide solutions to difficulties as they emerge;
(3) performance oriented - operates at an acceptable pace and is not overly burdensome to office personnel;
(4) non-intrusive - operates transparently; and
(5). functionally robust - provides ample coverage over essential areas.

This effort is an attempt to establish function- and context-memory residency within records, though the project has not yet determined the full requirements for identifying all contextual attributes—including not just the who, when, and why, but also the evolution of functions (i.e., why the transaction occurred in the first place and what action resulted: issues of evidence). As the project continues, these concerns are expected to be confronted in order to determine exactly what needs to be kept and how it will be captured. The results will be reflected in modified and expanded functional requirements that will be assessed by a group of government departments.

As with New York State, the chief role played by archivists is in establishing context for electronic records. As such, metadata capture and creation (through the protoype software) provide an operative strategy for preserving electronic records in a way that is intended to meet archival requirements.

United States National Archives and Records Administration[87]

Since 1989, NARA staff have been compiling a metadatabase—an entirely in-house initiative for capturing metadata on all of their holdings (including such items as agency history). They are currently expanding this project to include accessioned electronic records fonds. NARA envisions that eventually this metadatabase will become a formalized finding aid, though it is currently serving solely as an internal tool.

For electronic records systems, the project has been capturing information on logical file identity, record layouts, codebooks, and domain specifications. This metadata is used for reference purposes in order to inspect data after it has been accessioned by the Archives. Future goals include capture of metadata in order to allow "reactivation" of complex database structures as they originally existed within the agencies. NARA is itself keying in much of the metadata it now captures, since not all agencies present their descriptive documentation in electronic form. Policy discussions with agencies are being held in order to obtain their cooperation and to identify areas of mutual interest for this initiative, including life cycle management and functional requirements issues.

Responding to a recommendation from the Congressional Committee on Government Operations, which criticized NARA for not adequately ensuring the long-term preservation of electronic records systems, NARA noted that the preservation of these systems "may require intervention in the life cycle of a system to capture the data periodically...rather than waiting until the end of the life cycle which is defined by the agency...." Furthermore, NARA observed that to preserve these electronic information systems, agencies must manage them "in a way that adequately documents their policies, procedures, and activities; provide for the proper disposition of electronic records; and establish means for transferring valuable records to the National Archives."

NARA's Center for Electronic Records is suffering from inadequate resource allocation. As a consequence, the Center has commented that it has not been able to oversee the archival implications or management of electronic record-keeping in the federal government, and is lacking

staff expertise in areas such as data administration for handling such an undertaking. Understanding that it needs to develop tools that will enable it to preserve "electronic records with their specific 'record' characteristics intact," NARA has suggested that it needs to develop a data dictionary metadatabase in order to retain record and systems structure for data files and audit trail metadata for textual records. NARA views the proposed data element dictionary as a key finding aid for reference to data files.[88]

Of the four programmes surveyed, two are employing the metadata strategy to develop locator systems. Both Kentucky and New York State are creating ILSs documenting distributed systems across their respective bureaucracies. Kentucky has the added advantage of a forward-thinking information policy framework that facilitates metadata collection by the archives and ensures archival input into information systems planning and procurement. NARA is currently creating a solely in-house metadatabase resource. Canada is alone in actually influencing metadata creation within software applications, thus assisting the archival description process.

Conclusion

A metadata systems approach to the management of electronic records will become *the* basic strategy for archivists who must manage these systems. As computer systems reach the terminus of their useful lives, archivists must already have taken action if they are to have any hope of preserving the information resident on disks, drives, and tapes.[89] Otherwise, waiting to inventory the contents or the context of such information will overwhelm even the best-staffed institutions. Archivists must place the greater part of the descriptive burden for these systems on the systems themselves, through design intervention and via the users through policy implementation.

Margaret Hedstrom has noted that the absence of uniquely archival strategies with which to handle the range of records in modern electronic information systems has produced an enormous deficit, inhibiting the expansion of existing electronic records programmes and the development of new ones. She suggests that, since the profession cannot expect to influence the design of new information technologies, it should instead develop strategic interventions that place "more emphasis on designing systems and tools that support archival and information management objectives from the outset." Research by archivists will contribute positively to an agenda that addresses "concerns about access, retention, preservation, and future retrieval before these issues have been defined as too cumbersome, too expensive, or irrelevant."[90]

The University of Pittsburgh has just embarked upon a two-year NHPRC-funded research project that seeks to identify archival functional requirements for identifying, maintaining, and providing access to records in electronic records-keeping systems. Many of these functionalities will be captured and created as metadata, though the exact specifics of metadata capture and creation will only be developed through hypothesis and laboratory- and field-testing in assorted office and computerized environments.

What is clear is that post hoc description of electronic records systems will fail, given the amount of data likely to have to be evaluated. Provision for the creation and capture of accurate and comprehensive metadata in the systems design stage provides the archival community with an operative strategy for capturing the crucial contextual information that provides an understanding of electronic records and the records-keeping systems in which they reside. Archivists can take advantage of the relevant metadata that these systems already produce and conduct research on the types of metadata that they will want created in order to ensure the identification and preservation of understandable and interpretable records. This objective inherently requires a proactive strategy, new technological skills, a multidisciplinary outlook, and enhanced political leverage within parent organizations.

Notes

* The author would like to thank Margaret Hedstrom, John McDonald, Charles Robb, and Kenneth Thibodeau for agreeing to be interviewed and for sending valuable grey literature; David Bearman for commenting on an earlier draft of this paper, providing access to his unique library, and sharing many hours of rich discussion on electronic records issues; and Richard Cox for commenting on an earlier draft of this paper and providing the opportunity for an abbreviated version of it to be read at the Spring 1993 meeting of the Midwestern Archivists Conference in Chicago.

1 Jerry M. Rosenberg, *Dictionary of Artificial Intelligence and Robotics* (New York, 1986), p. 117.

2 This definition has been used in three formal archival glossaries: see United Nations. Advisory Commission for the Co-ordination of Information Systems, *Management of Electronic Records: Issues and Guidelines* (New York, 1990), p. 167; Charles Dollar, *Archival Theory and Information Technologies: The Impact of Information Technologies on Archival Principles and Methods* (Macerata, Italy, 1992), p. 87; and Lewis J. Bellardo and Lynn Lady Bellardo, *A Glossary for Archivists, Manuscript Curators, and Records Managers* (Chicago, 1992), p. 22. They are derivative of the definition offered in Belkis W. Leong-Hong and Bernard K. Plagman, *Data Dictionary/Directory Systems: Administration, Implementation and Usage* (New York, 1982), pp. 320-21.

3 Terry Cook, "Easy to Byte, Harder to Chew: The Second Generation of Electronic Records Archives," *Archivaria* 33 (Winter 1991-92), pp. 205-206.

4 United Nations. Advisory Committee for the Co-ordination of Information Systems, *Strategic Issues for Electronic Records Management: Towards Open Systems Interconnection* (New York, 1992), p. 9.

5 Judi Moline, *Recommendations for Document Transfer Standards and Their Integration into National Archives Policy.* Attachment C: "Framework for the Exchange and Preservation of Electronic Records. Report Prepared for the National Archives and Records Administration by the Systems and Software Technology Division of the National Institute of Standards and Technology," January 1989, p. 4.

6 Terry Cook, "Easy to Byte, Harder to Chew,", pp. 205-6.

7 Margaret Hedstrom, "Understanding Electronic Incunabula: A Framework for Research on Electronic Records," *American Archivist* 54 (Summer 1991), pp. 338, 350.

8 *Archival Administration in the Electronic Information Age: An Advanced Institute for Government Archivists* (15 August 1989), p. 16. Organized and Conducted by the School of Library and Information Science, University of Pittsburgh; co-sponsored by the National Association of Government Archives and Records Administrators, with support from the Council on Library Resources.

9 Michael Buckland, *Information and Information Systems* (New York, 1991), pp. 69-71.

10 Terry Cook, "Easy to Byte, Harder to Chew,", pp. 206-7.

11 Ibid., p. 207.

12 Margaret Hedstrom, "New Appraisal Techniques: The Effect of Theory on Practice," *Provenance* 7 (Fall 1989), pp. 20-21.

13 For a detailed discussion of this issue see David Bearman, *Archival Methods* (Pittsburgh, 1989), pp. 6-16.

14 Terry Cook, "Mind Over Matter: Towards a New Theory of Archival Appraisal," in Barbara L. Craig, ed., *The Archival Imagination: Essays in Honour of Hugh A. Taylor* (Ottawa, 1992), pp. 42-43.

15 NARA's Center for Electronic Records currently employs forty-two full-time and eight part-time employees. However, a recent discussion of its operations has commented that "[t]he growth in the Center's holdings and workload has far outstripped the growth in its staff": Status report presented to a joint meeting between NAGARA's Committee of Information Technology and SAA's Committee on Automated Records and Techniques, 23 April 1993.

16 Committee on the Records of Government, *Report* (Washington, 1985), pp. 24, 26, 78, and 89. For a discussion of the magnitude of the growth of federal computing, see James Gregory Bradsher, "A Brief History of the Growth of Federal Government Records, Archives, and Information, 1789-1985," *Government Publications Review* 13 (1986), pp. 497-501.

17 Ibid., pp. 27-34.

18 National Archives and Records Administration. Office of Records Administration, *Managing Electronic Records* (Washington, 1990), pp. 26-27.

19 Presentation by Kenneth Thibodeau at the "Experiment in Interdisciplinary Appraisal: The NAPA II Project" Session: 1992 Annual Meeting of the National Association of Government Archives and Records Administrators, Washington, DC, 17 July 1992.

20 National Academy of Public Administration, *The Archives of the Future: Archival Strategies for the Treatment of Electronic Databases* (Washington, 1991), p. vi.

21 Ibid., pp. 27-28.

22 Richard J. Cox, "Archivists, Electronic Records, and the Modern Information Age: Re-Examining Archival Institutions and Education in the United States, with Special Attention to State Archives and State Archivists" (Ph.D. dissertation, University of Pittsburgh, 1992), p. 19.
23 Ibid., p. 254. For a fuller discussion of the paucity of relevant electronic records courses and the general inadequacy of the existing archival educational structure, see pages 236-85.
24 Ibid., p. 73.
25 Hedstrom, "Understanding Electronic Incunabula," p. 336.
26 For example, see Hedstrom, "New Appraisal Techniques," p. 16; Michael Miller, "Is the Past Prologue? Appraisal and the New Technologies," in David Bearman, ed., *Archival Management of Electronic Records* (Pittsburgh, 1991), pp. 40-41; Charles M. Dollar, *Archival Theory and Information Technologies*, p. 74; and, United Nations Advisory Committee for the Co-ordination of Information Systems, *Management of Electronic Records: Issues and Guidelines*, pp. 23 and 39; Hamza Kandur, "Management of Electronic Records" (Ph.D. dissertation, University of London, 1992), pp. 31-37.
27 United Nations, *Strategic Issues for Electronic Records Management* 6. The traditional archival life cycle, employed by the United States National Archives and Records Administration, contains the following stages:
(1) Creating Documents and Scheduling
(2) Disseminating
(3) Filing (Retrieve, Re-Use, Destroy)
(4) Transferring (to Archives)
(5) Preserving
(6) Retrieving (for Use)
Source: Judi Moline, *Recommendations for Document Transfer Standards and Their Integration into National Archives Policy*, p. 6.
28 Archival functional requirements are those criteria which must be met in order to render records understandable and usable for archival purposes. This includes data capture, storage, preservation, and access.
29 Belkis W. Leong-Hong and Bernard K. Plagman, *Data Dictionary/Directory Systems: Administration, Implementation and Usage* (New York, 1982), p. 18.
30 Ibid., p. 165.
31 Gary B. Shelly, Thomas J. Cashman, Judy Adamski, and Joseph J. Adamski, *Systems Analysis and Design* (Boston, 1991), 4.22.
32 Leong-Hong and Plagman, p. 18.
33 Rom Narayan, *Data Dictionary: Implementation, Use, and Maintenance* (Englewood Cliffs, NJ, 1988), p. 7.
34 Ibid., pp. 7-8.
35 Ibid., p. 8.
36 Peter Sigmond, "Form, Function and Archival Value," *Archivaria* 33 (Winter 1991-92), pp. 141-47.
37 Bruno Delmas, "Archival Science and Information Technologies," In Angelika Menne-Haritz, ed., *Information Handling in Offices and Archives* (Munich, 1993), p. 173.
38 Shelly, et al., *Systems Analysis and Design*, 4.1-4.28.
39 Ibid., 4.2. It is described in the text as "data input to a process, and the process transforms this data to produce output data of a different content or form."
40 This example is meant to be representational of the type of formalized document flows that are present in organizations.
41 JoAnne Yates, "Internal Communication Systems in American Business Structures: A Framework to Aid Appraisal," *American Archivist* 48 (Spring 1985), pp. 141-58.
42 W.H. Inmon, *Building the Data Warehouse* (Wellesley, MA, 1992), pp. 156-57, 196-97, 253.
43 Simon Holloway, *Data Administration* (Aldershot, 1988), p 7.
44 Ibid., pp. 50-51.
45 Ibid., pp. 54.
46 David Bearman, *Archival Methods* (Pittsburgh, 1989), p. 37.
47 Ibid., p. 36.
48 Ibid., pp. 26-27.
49 Richard M. Kesner, "The Changing Face of Office Documentation: Electronic/Optical Information Technologies (IT)," In Angelika Menne-Haritz, ed., *Information Handling in Offices and Archives*, pp. 114-115; Hamza Kandur, "Management of Electronic Records," pp. 128, 138, argues this point as well.
50 National Historic Publications and Records Commission, *Electronic Records Issues: A Report to the Commission, Commission Reports and Papers* 4 (March 1990), p. 8.

51 National Historical Publications and Records Commission, *Research Issues in Electronic Records* (St. Paul, 1991), pp. 13-14.
52 National Historical Publications and Records Commission, "Electronic Grant Guidelines: Draft" (4 November 1992).
53 Described as

> data elements (often in standardized formats) that effectively 'describe' one or more characteristics of the information and/or records in the system. They might cover a file's physical characteristics (what kind of hardware it runs on, how many separate records exist in the file and how much room each occupies); when, why and by whom the file was created or updated; or instruction for transfer or erasure of discrete pieces of information after a specified period of time.

54 "Report of the Working Group on Standards for Archival Description," *American Archivist* 52 (Fall 1989), p. 442.
55 Prepared by David Bearman for the Advisory Committee for the Co-ordination of Information Systems (ACCIS), (New York, 1990).
56 Ibid., p. 25.
57 Ibid., pp. 20-25.
58 Ibid., pp. 37, 56.
59 Ibid., pp. 58, 64-65.
60 David Bearman, "Archival Data Management to Achieve Organizational Accountability for Electronic Records," version of paper presented at the Fifth Annual Congress of the Portuguese Association for History and Computing, Lisbon, Portugal, 2 July 1992; and "New Models for Management of Electronic Records by Archives," *Cadernos De Biblioteconomia, Arquivistica E Documentacao* 2 (1992), pp. 61-70.
61 United Nations, *Strategic Issues for Electronic Records Management*, pp. 6-11.
62 Ibid., pp. 97-98.
63 Hamza Kandur, "Management of Electronic Records," pp. 102, 132-36.
64 Ibid., pp. 142-44. The wisdom of actually transferring the data file into an archive has been a subject of debate. While it is believed that the counter-argument against accessioning electronic data systems is sound, Kandur's model directory is proposed in the belief that its utility goes beyond whether or not systems are accessioned.
65 Ibid., p. 177.
66 Dollar, *Archival Theory and Information Technologies*, p. 18.
67 Ibid., pp. 45-51.
68 Ibid., p. 108. Victoria Irons Walch, in a review of technical standards applicable to archivists, noted that the IRDS "offers one of the most archivally intriguing standards in terms of exploiting the ability of automated systems to 'describe' themselves and for providing the finding aids of the future": "The Role of Standards in the Archival Management of Electronic Records," *American Archivist* 53 (Winter 1990), p. 41.
69 Ibid., p. 59.
70 Ibid., p. 62.
71 Ibid., pp. 62-69.
72 Ibid., pp. 81-82.
73 Margaret Hedstrom, "Descriptive Practices for Electronic Records: Deciding What is Essential and Imagining What is Possible," paper presented at the Annual Conference of the Association of Canadian Archivists, Montréal, Québec, 12 September 1992; it is published elsewhere in this issue of *Archivaria*.
74 Ibid., p. 5.
75 Ibid., pp. 6-7.
76 Ibid., pp. 8-9.
77 Menne-Haritz, ed., *Information Handling in Offices and Archives*.
78 Ibid., pp. 9-26.
79 Richard E. Barry, "Getting it Right: Managing Organizations in a Runaway Electronic Age," in Menne-Haritz, ed., *Information Handling in Offices and Archives*, pp. 50-51.
80 David Bearman, "Archival Principles and the Electronic Office," in Menne-Haritz, ed., *Information Handling in Offices and Archives*, pp. 177-83.
81 Ibid., pp. 183-93.
82 Richard M. Kesner, "Automated Information Management: Is There a Role for the Archivist in the Office of the Future," *Archivaria* 19 (Winter 1984-85), pp. 162-72.

83 Margaret Cross Norton, "The Archivist and Records Management," in Thornton W. Mitchell, ed., *Norton on Archives: The Writings of Margaret Cross Norton on Archival and Records Management* (Carbondale, IL, 1975), p. 248.

84 Discussion based on telephone interview with Chuck Robb, Kentucky Department for Libraries and Archives, 16 October 1992; Richard N. Belding and Larry G. Forston, "Kentucky's Machine-Readable Challenge," *Provenance* 3 (Spring 1985), pp. 58-69; Chuck Robb, "Information Resource Management in Kentucky State Government," *Archives & Museum Informatics* 5 (Winter 1991), pp. 2-4; and "Electronic Records Program Summary," Paper presented to a joint meeting of NAGARA's Committee on Information Technology and SAA's Committee on Automated Records and Techniques, 23 April 1993.

85 Discussion based on telephone interview with Margaret Hedstrom, New York State Archives and Records Administration, 22 October 1992; New York State Forum for Information Resource Management reports: "New York State Sourcebook Pilot Project: Final Evaluation Report" (March 1992); "New York State Sourcebook Pilot Project: A Metadata Approach to Information Management" (March 1992); and "From The Sourcebook to a Government-wide Information Locator System: Next Steps for Sourcebook Expansion" (March 1992).

86 Discussion based on telephone interview with John McDonald, National Archives of Canada, 26 October 1992; "Strategic Framework for the Information Management Standards and Practices Division" (April 1991); "Strategic Priorities for the Information Management Standards and Practices Division" (April 1991); "Strategic Goals for the Information Management Standards and Practices Division" (17 July 1991); and "The IMOSA (Information Management and Office Systems Advancement) Project: Phase I Report" (1991).

87 Discussion based on telephone interview with Kenneth Thibodeau, Center for Electronic Records, National Archives and Records Administration, 22 October 1992; and National Archives and Records Administration, "Management, Preservation and Access for Electronic Records With Enduring Value: A Response to Recommendations in 'Taking a Byte Out of History: The Archival Preservation of Federal Computer Records,'" 1 July 1991.

88 National Archives and Records Administration, Center for Electronic Records, "Strategic Analysis: The Implications of Electronic Records on the Programs, Activities and Responsibilities of the National Archives and Records Administration" (May 1989).

89 A front page article in the *New York Times* (14 April 1993) noted that in the United States, businesses and individuals are getting rid of enormous numbers of computers. Presumably, masses of data on these devices are suffering the same fate as well.

90 Hedstrom, "Understanding Electronic Incunabula," pp. 336-37.