

Rivista quadrimestrale

Roberto CERRI

Elisa PANICUCCI

Maria Bruni BALDACCI, Gabriele BEZZI, Marina BROGI, Euride FREGNI, Bruno GIALLUCA, Bruna GODONE, Roberto GRASSI, Roberto GUARASCI, Mariella GUERCIO, Maria Teresa INGROSSO, Ilaria PESCHINI, Claudia SALMINI, Sandra SAMBO, Michela SESSA, Maria SIMI, Giuliano TERZUOLI, Stefano VITALI, Gabriele ZACCAGNINI.

Elisa PANICUCCI

"ARCHIVI & COMPUTER"

Archivio Storico Comunale, Loggiati di San Domenico, 3 - 56027 San Miniato (Pi)
Tel. 0571 - 418381; Telefax 0571 - 406233; E-Mail: archilab@comune.san-miniato.pi.it
<http://www.archi-lab.it>

Abbonamento «persona fisica»:	Euro 31,00
Abbonamento «biblioteche, istituzioni, enti pubblici»:	Euro 36,10
Abbonamento estero:	Euro 51,60
Prezzo del singolo fascicolo:	Euro 12,90
Sconto per librerie:	10%
Sconto per soci Archilab:	30%

Le richieste di abbonamento devono pervenire alla **Carocci Editore S.p.A.**, via Sardegna, 50 - 00187 Roma, tel. 06-42014260, fax 06-42747931, e-mail: riviste@carocci.it
I relativi importi devono essere versati sul conto corrente postale n. 77228005 intestato a Carocci Editore S.p.A., via Sardegna, 50 - 00187 Roma.

Per il momento è attivabile l'abbonamento 2002 alla sola versione cartacea. La redazione sta valutando l'opportunità di mantenere per il 2002 la versione elettronica sperimentata nel 2001. Nel caso che la versione elettronica non venga riattivata, sarà prevista una forma di rimborso per chi avesse già versato la relativa quota di abbonamento.

Sono disponibili i fascicoli e le annate 1991-2001 (un fascicolo Euro 12,90; un'annata Euro 51,60; sconto per librerie 10%; sconto per soci 30%). Le richieste devono pervenire ad Archilab, c/o Archivio Storico Comunale, Loggiati di San Domenico, 3 - 56027 San Miniato (Pi) - Tel. 0571 - 418381; Telefax 0571 - 406233; E-Mail: archilab@comune.san-miniato.pi.it e i relativi importi devono essere versati sul conto corrente postale n. 12252532 intestato ad Archilab, oppure effettuando un bonifico sul conto corrente bancario n. 10009004 intestato ad Archilab - Cassa di Risparmio di San Miniato - Ag. San Miniato - ABI 6300 - CAB 71150.

Autorizzazione del Tribunale di Pisa n. 8 del 23.03.1991.
Quadrimestrale - Spedizione in A.P. - 70% - Filiale di Pisa.
Editore: Archilab, Loggiati di S. Domenico, 3, 56027 San Miniato (Pi).
Stampa: Litografia Varo, via Carducci 51, 56010 Ghezzano (Pi).

Il quadrimestre 2001 - Finito di stampare nel mese di gennaio 2002

ISSN 1121-2462

Archivi & Computer

Anno XI

Fascicolo 2/2001

		p. 89
Reagan W. MOORE Kenneth THIBODEAU	<i>Knowledge-based Persistent Archives Digital Preservation Techniques: Evaluating the Options</i>	p. 91 p. 101
Anne J. GILLILAND- SWETLAND	<i>The Potential of Markup Languages to Support Descriptive Access to Electronic Records: The EAD Stan- dard</i>	p. 110
Barbara VALLI - Francesco CATTANEO	<i>Dai documenti alla scena. Sul tenta- tivo di portare in teatro gli archivi e le loro storie</i>	p. 122
Guido MONTANARI - Francesca PINO	<i>Un sistema informativo integrato per gli archivi storici: le funzionalità di GEA-IntesaBci per l'utente Internet e per l'archivista</i>	p. 131
Lucia NARDI Cecilia POGGETTI	<i>L'ANAI e la professione archivistica Archivi aperti al pubblico</i>	p. 144 p. 145
Paola GADDO - Francesca ROCCHETTI	<i>Progetto C.I.R.C.E. (Catalogo Infor- matico delle Riviste Culturali Euro- pee)</i>	p. 148
Sara GUIATI	<i>«Le management de l'archive»: ar- chivio ed archiviazione al centro di un'opera pluridisciplinare</i>	p. 151

Anne J. GILLILAND-SWETLAND

The Potential of Markup Languages to Support Descriptive Access to Electronic Records: The EAD Standard

L'articolo esamina le potenzialità di Encoded Archival Description (EAD), recentemente adottato come standard di descrizione archivistica negli USA, ai fini di fornire accesso descrittivo on line ai documenti elettronici. L'autrice mette in evidenza le complesse relazioni che intercorrono tra metadati che sono parte integrante del documento elettronico e metadati che descrivono il documento medesimo. Le caratteristiche di EAD sono illustrate anche attraverso il confronto con altre iniziative descrittive che utilizzano linguaggi di marcatura.

Introduction

There has been a considerable amount of political and professional rhetoric in the United States and Europe, stemming from unprecedented developments over the past decade in technologies supporting the World Wide Web, about developing online access to unpublished information resources – including archival holdings. This rhetoric has resulted in the establishment of research and development agendas by major government funding agencies, private foundations, industry, and professional institutions and associations¹. A number of major initiatives have resulted from the availability of this funding. As they relate to archival concerns, these initiatives can be grouped into three primary domains of activity:

1. the development of archival standards that support online access to archival descriptions (Encoded Archival Description (EAD) being the most prominent recent example);
2. the development of archival information systems such as American Memory

¹ For example, the National Science Foundation, the National Endowment for the Humanities, and the National Historical Publications and Records Commission in the United States, and the Fifth Framework and the Joint Information Systems Committee in Europe; national archives and libraries in many countries; and descriptive standards groups within professional associations.

- at the Library of Congress and the Online Archive of California (to cite two American examples) that provide not only online descriptions but also digitized copies of selected archival holdings; and
3. research projects addressing the archival management of records that are «born digital», that is, of electronic records (for example, the Recordkeeping Functional Requirements Project at the University of Pittsburgh and the International Research on Permanent Authentic Records in Electronic Systems Project (InterPARES))².

While there has been considerable dialog and overlap between archivists involved with the first two of these areas, until recently archivists grappling with the challenges of creating and preserving electronic records have not been integrally engaged in broader initiatives to standardize and enhance description for online access, nor to provide online access to electronic records through archival information systems or digital libraries. A major exception has been the Australia, where the Recordkeeping Metadata Schema (RKMS) provides an integrated approach to recordkeeping metadata and systems development for records, regardless of their format or institutional contexts; and a mapping of that metadata into metadata schema used by other information communities. The focus for RKMS is upon how the record can be reconstructed and retain its meaning across time and user domains through an identification of the metadata associated with records and recordkeeping processes and agents³.

In North America, electronic records management evolved to some extent as an area apart from the mainstream of the archival profession. Its immediate concerns have been on creating, identifying, and accessioning electronic records. In the 1970s and 1980s, electronic records, or «machine-readable records» as they were initially termed, tended to be managed as software-independent datafiles. More recently, as electronic records have taken on more complex functionality, there has been an increased awareness of the need to preserve their value as legal and organizational evidence. As a result, archivists are now engaged with researchers from computer science, digital library development, and preservation in several projects to identify how to preserve authentic electronic records with their functionality intact. One of the most prominent of such projects is that of the National Archives and Records Administration and the San Diego Supercomputer

² A.J. Gilliland-Swetland and P. Eppard, *Preserving the Authenticity of Contingent Digital Objects: The InterPARES Project*, «D-Lib Magazine», 6 no. 7 (2000). Available at: <http://www.dlib.org/dlib/july00/eppard/07eppard.html> (16 October, 2000); InterPARES Website available at <http://www.interpares.org> (16 October, 2000).

³ RKMS does this through the provision of a standardized set of structured or semi-structured recordkeeping metadata elements; a framework for developing recordkeeping metadata sets in different contexts; and a framework for mapping recordkeeping metadata sets to establish equivalences and correspondences that can provide the basis for semi-automated translation between metadata sets. See S. McKemmish, G. Acland, N. Ward and B. Reed, *Describing Records in Context in the Continuum: The Australian Recordkeeping Metadata Schema*, «Archivaria», 48 (1999), pp. 3-42.

Center to employ Extensible Markup Language (XML) in the development of persistent archives through a process termed Persistent Object Preservation. This project explicitly focuses upon meeting archival integrity requirements rather than merely addressing issues of technological obsolescence. The project seeks to transform electronic records automatically by expressing archival requirements as formal models that are embedded in the preservation infrastructure, and analysing and then tagging in XML the significant elements and properties of records needing to be preserved in order that they can subsequently be reconstructed independent of the software-hardware infrastructure on which they were created.

The heightened archival concern for evidence requires a more detailed understanding of what are the characteristics of an authentic record in and over time. It also requires close analysis of the intellectual rationales underlying archival description in terms of how it contributes to ensuring and demonstrating the authenticity and continued accessibility of preserved records. This concern is evident in both the RKMS and the Persistent Object Preservation initiatives. It is also at the heart of the InterPARES Project. The goal of the InterPARES Project is to develop the theoretical and methodological knowledge essential for the permanent preservation of records generated electronically, and, on the basis of this knowledge, to formulate model policies, strategies, and standards capable of ensuring their preservation for use by archivists and other communities who need to retain and use trustworthy electronic records over indefinite periods of time. The work of the project has been broken down into four research domains: conceptual requirements for preserving authentic electronic records; appraisal criteria and methodology for authentic electronic records; methodologies for preserving authentic electronic records; and frameworks for developing policies, strategies and standards.

Archives play a key and often overlooked role in establishing and demonstrating the authenticity of any record, regardless of its form, through archival description. In contrast to the key purposes of bibliographic description which are to manage a physical information object as well as to facilitate its intellectual retrieval and use, archival description must address that object not only as information, but also as evidence. As a result, archival description must not only describe the content of a fonds or record group, it must also describe the circumstances of its creation, its chain of custody, its relationships to other records generated by the same activity, and the impact upon the aggregation of records of any processing or preservation activity in ways that are and remain meaningful to different kinds of users over time. Archival description, therefore, has three primary roles. Firstly, it serves as a tool that meets the needs of the archival records being described by authenticating and documenting them. Secondly, it is an archival management tool. Thirdly, it is an information discovery and retrieval tool for making the evidence and information contained in archives available to and comprehensible by archivists and users alike.

Today there is a growing convergence within the archival profession including electronic records archivists, as well as of other professional and disciplinary domains around issues relating to description and other kinds of metadata. This convergence arises largely out of the development of new metadata schema, standards and technological capabilities such as Standard Generalised Markup Language (SGML), XML, and Resource Description Format (RDF) that provide structures, frameworks, and crosswalks⁴ for formalizing and bridging diverse data types (such as image or geospatial data), metadata semantics, and professional practices⁵.

Describing Electronic Records

Ironically, in a world of increasing online access to primary information resources, many of which first require digitization, electronic records that by their very nature are «born digital,» are proving to be among the most intractable of information resources in terms of providing even basic descriptive access. This intractability reflects inherent technical problems with the diverse formats in which electronic records are created and may need to be maintained and used. Equally, it reflects how the enormous volume of electronic records requiring archival processing by comparatively small staffs together with data archiving practices originally adopted from the social sciences data archives community have led to idiosyncratic archival description and an over-dependence upon the metadata generated by the creator of the records. In essence, archival description of electronic records is unstandardised, consisting of components such as high level summaries of data, reports on quality and accuracy of data, scanned or Portable Document Format (PDF) versions of codebooks and data dictionaries, and customized subject indexes and data extracts.

While the current state of description for electronic records is certainly understandable, it is, nevertheless, deficient in several respects:

- There has been insufficient analysis of what is the actual nature of electronic records. In particular, there needs to be more examination of the relationship between data content and the metadata that provide and document its creative and documentary context and structure, and of the various ways in which aspects of data and metadata in complex systems such as databases might come

⁴ A crosswalk is a chart or table that represents the mapping of fields or data elements in one metadata standard to fields or data elements in other standards that have the same function or meaning. Crosswalks support the ability to search transparently heterogeneous databases as a single database (semantic interoperability) and to convert data from one metadata standard to another.

⁵ See A.J. Gilliland-Swetland, *Enduring Paradigm, New Opportunities: The Value of the Archival Perspective in the Digital Environment*, Washington, D.C.: Council on Library and Information Resources, 2000.

together to form the intellectual and functional construct that is a record. Often one of the most difficult aspects of working with electronic records is to be able to identify and then describe, in the absence of a tangible document, the parameters of that intellectual construct.

- Metadata generated by records creators has been viewed as sufficient substitute for archival description. For example, in 1993, Margaret Hedstrom, a prominent United States electronic records archivist and researcher, proposed that management of metadata provide an alternative strategy to current descriptive practices in order to support the «need to identify, gain access, understand the meaning, interpret the content, determine authenticity, and manage electronic records to ensure continuing access»⁶. Subsequently, several research projects have generated metadata specifications for electronic records, most notably the Pittsburgh Project and related implementation projects such as the Indiana University Electronic Records Project. With the exception of the Australian RKMS project, there has been almost no discussion, however, of the value-added role that archival description should play in terms of ensuring and documenting authenticity, and making the records meaningful to users across time and domains⁷.
- There has been little emphasis on establishing the documentary relationships between electronic records and paper records created by the same activity. Additionally, lack of standardization and use of non-archival descriptive practices has made it difficult to integrate descriptions of electronic records with standardized descriptive metadata created by archivists and other information, industry, and research communities for traditional records of the same fonds. For example, in the mid-1980s, when United States archivists looked to the use of MARC formats to standardise and integrate high-level archival description into wider online information systems, electronic records archivists in some major repositories opted to employ the MARC Machine-Readable Data Format (MRDF) rather than the MARC Archives and Manuscripts Control Format (AMC) that was developed for the collective description of archival and manuscript materials. In effect, such an approach treated electronic records as a special format with distinct descriptive needs, rather than as components of wider archival aggregations.
- Because management of electronic records has generally been viewed by the

⁶ M. Hedstrom, *Descriptive Practices for Electronic Records: Deciding What is Essential and Imagining What is Possible*, «Archivaria» 36 (Autumn 1996), p. 53.

⁷ D. Bearman and K. Sochats, *Metadata Requirements For Evidence*, 1996, Available: <http://www.lis.pitt.edu/~nhprc/BACartic.html> (October 17, 2000); P.C. Bantin, *Developing a Strategy for Managing Electronic Records: The Findings of the Indiana University Electronic Records Project*, «American Archivist» 61 (1998), pp. 328-64; P.C. Bantin, *The Indiana University Electronic Records Project Revisited*, «American Archivist» 62 (1999), pp. 153-163; S. McKemmish, G. Acland and B. Reed, *Towards a Framework for Standardising Recordkeeping Metadata: The Australian Recordkeeping Metadata Schema*, «Records Management Journal», 9 (1999), pp. 177-202.

rest of the archival profession as an area that requires distinct technical expertise, developments in archival description such as EAD have progressed without being strongly informed by the descriptive needs of electronic records.

It is useful at this point to define more closely what is meant here by metadata, since the term is understood differently by different communities. Metadata refers to a range of structured or semi-structured data about data that are critical to the development of effective, authoritative, interoperable, scalable, and preservable information and record-keeping systems. Until the mid-1990s, metadata was a term most prevalently used by communities involved with the management and interoperability of geospatial data, and with data management and systems design and maintenance in general. For these communities, metadata referred to a suite of industry or disciplinary standards as well as additional internal and external documentation and other data necessary for the identification, representation, inter-operability, technical management, performance, and use of data contained in an information system. For archivists, metadata refers to the value-added information, such as EAD, that they create in order to identify, authenticate, arrange, reconstruct, describe, preserve, and otherwise enhance access to their holdings.

In contemplating the role of metadata in the description of electronic records, several questions come to mind:

- Which metadata are part of the record, which are about the record, and which are neither but are required to preserve or reconstruct the technological context of the record? And of all these types of metadata, which must be captured as part of archival description?
- How can the trustworthiness of these metadata be determined in terms of quality and completeness in and over time?
- Are there descriptive needs of electronic records that might be different from those of other types of records? If so, what are they and how should they best be addressed within a recordkeeping, archival management, or information management framework?
- Can appropriate aspects of the metadata generated by the creator of the electronic record somehow be translated or mapped automatically into a standardized description for archival records?
- Can the structure and documentary contexts of electronic records be analyzed automatically to generate specific components of a standardized description for electronic records?
- Which kinds of contextual documentation do electronic records require in order to be understood and can a metadata infrastructure facilitate links to that documentation online?
- How can the links between records and metadata retain their referential integrity over time in the face of systems obsolescence, data migration, and evolution of metadata schema?

- Which kinds of description do users need in order to be able to identify relevant electronic records online? What do users need to be able to use electronic records disseminated online efficiently and effectively?

Encoded Archival Description

In the face of such questions, therefore, how might SGML-based markup languages such as XML, and Encoded Archival Description in particular enhance electronic records description? SGML (Standard Generalized Mark-up Language) is an international standard (ISO 8879) that provides a grammar for formally defining, specifying, and creating hierarchical and object-oriented structural standards or document type definitions (DTDs) for digital documents that can be delivered, displayed, linked, and manipulated in a platform-independent manner. XML is a simplified subset of SGML designed for use on the Internet. XML supports a wide variety of applications and provides an easy way for creators of documents easily to develop a concise, semi-structured data model to create and validate those documents in a predictable manner.

Regardless of the implementation, there are many benefits to using SGML-based mark-up to develop predictable document structures:

- Document structures are consistent, thus facilitating searching across documents at the same or at multiple sites, as well as validation of individual documents.
- SGML parsing of a DTD provides one means for assessing the reliability of documents.
- Hierarchical and object-oriented structures potentially facilitate multiple retrieval mechanisms at different levels of granularity (e.g., at both collection and item-level) for both browsing and retrieval. In general, XML documents are much «smarter» than HTML documents because of their predictable structures.
- Marked-up documents are easily updated or otherwise edited.
- Marked-up documents can provide users with a visual representation of the relationships between distributed and multi-level data elements that they contain.
- DTDs provide a way to incorporate non-textual media and unstructured or semi-structured information objects such as images and sound.
- DTDs are platform independent and facilitate long-term preservation because their migration paths are standardised.

Simply defined, EAD is a Document Type Definition (DTD) developed using SGML that identifies and standardises archival descriptive and administrative data elements and the relationships between those elements so that predictably structured, hierarchical, archival descriptions that can be created and disseminated on the World Wide Web. That description is most commonly in the form of an archival finding aid, but the DTD is flexible enough to accommodate various oth-

er types of archival descriptive tools.

However, the real potential of EAD is to be much more than a structure used to create a digital representation of a two-dimensional paper finding aid. The hierarchical nature of EAD, its explicit delineation of each data element, and its adherence to standardised metadata conventions and protocols provide it with the potential to function as a multi-dimensional metadata infrastructure. This infrastructure is able to interface with other predictable metadata schema, as well as to provide maximum flexibility in describing a diversity of record types in ways that reflect the organic nature of archives. With such an infrastructure, archivists and software developers have the capabilities and incentives to design a range of archival information systems that fundamentally re-conceptualize how access to archives is provided. These archival information systems would not only contain the kinds of archival description found today in finding aids, but also digitized versions of archival materials, full-text of ancillary materials, relevant linkages to other online archival and bibliographic information systems, and actual electronic records and the technical documentation required to use them»⁸.

In such information systems, however, EAD would not be the only metadata schema at work, and one of the powerful aspects of EAD is its ability to interface or interoperate with other metadata schema and SGML-based implementations. EAD is fully XML-compliant, meaning not only that EAD-encoded descriptions can be more extensively searched and manipulated online as the Web increasingly supports XML, but also that electronic records technical documentation, such as database models, workflow rules, and technical drawings can be integrated with the archival descriptions in ways not previously possible in a more manual environment. Similarly, EAD can interface with descriptive metadata created in MARC because of metadata mapping between the two standards. With the recent release of XMLMARC software, this mapping will become only easier. EAD also shares header data elements with the Text Encoding and Interchange (TEI) DTD, a DTD that facilitates the development of digital versions of scholarly texts.

Using EAD to Describe Electronic Records

EAD is currently in its first full release (Version 1.0). DTDs are dynamic, rather than static constructs, however, and the EAD DTD will continue to be extended to accommodate new technological capabilities and metadata schema, as

⁸ A.J. Gilliland-Swetland, *Popularizing the Finding Aid: Exploiting EAD to Enhance Online Browsing and Retrieval in Archival Information Systems by Diverse User Groups*, «Journal of Internet Cataloging» 4 nos. 1/2 (2000) (in press); A.J. Gilliland-Swetland, *Health Sciences Documentation and Networked Hypermedia: An Integrative Approach*, «Archivaria», 41 (1995), pp. 41-56.

well at to be refined based on evaluative feedback from archivists and users. While the descriptive needs of electronic records were not integrally addressed by the current DTD, EAD nevertheless offers advantages for electronic records description.

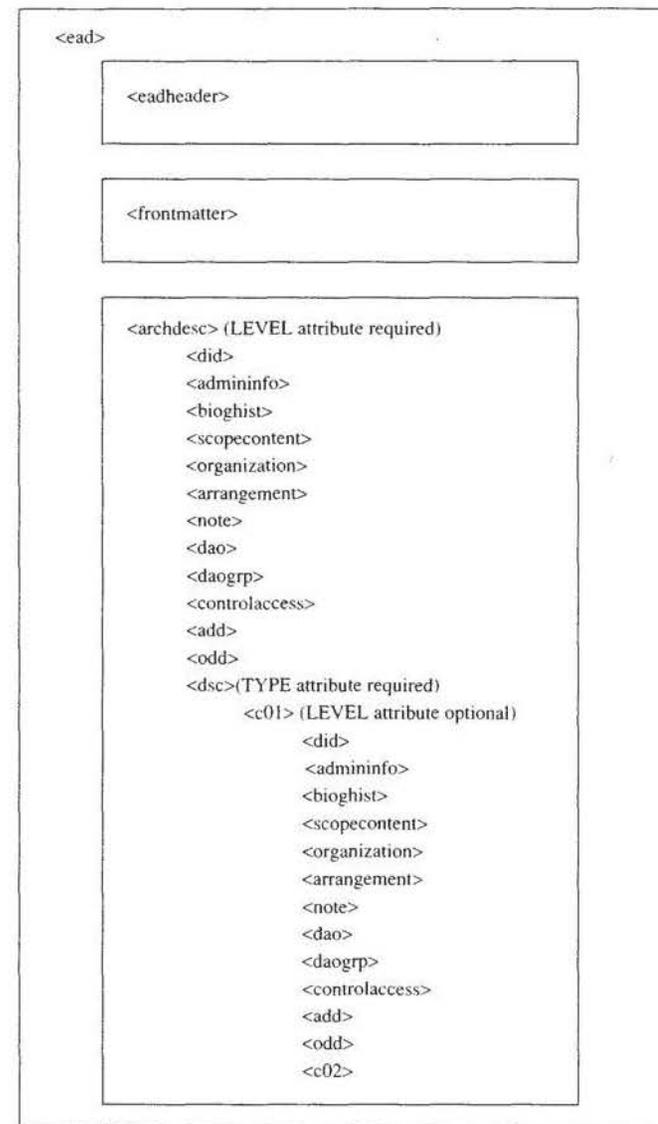
EAD, while it is a data structure and not a data content standard, works to standardize idiosyncratic descriptive practices. Electronic records descriptive practices are some of the most idiosyncratic in the field in part because there is such diversity of types of electronic records and programs, and in part because electronic records description is rarely taught in archival education programs, and is primarily learned as institution-specific practices «on the job.» Descriptive records tend to comprise examples of descriptions of datafiles, rather than complete descriptions, together with user guides of documentation packages⁹. Using EAD would also integrate electronic records management into the mainstream of archival activities, treating the records as records, rather than as instances of special formats. Moreover, through collective description, as well as EAD elements such as <separatedmaterial> and <relatedmaterial>, all records created by the same activity will be treated as an intellectual whole, regardless of whether they are paper, electronic, or some other medium (see Figure 1).

Compared to finding aids for paper records, electronic records descriptions can be quite flat, consisting mostly of summary information and little descriptive hierarchy. Original order is often treated as incidental since the contents of datafiles can be arranged in multiple ways. However, electronic records users may still wish to have access at the level of individual records, files, or even data elements. The hierarchy built into EAD has the potential to support this kind of granularity of access, although commercial software that is currently available has yet to address much of this potential. Technical documentation accompanying the electronic records can also be linked in electronic form to the EAD description through elements such as <archref>, <odd> (other descriptive data) and <add> (adjunct descriptive data). If this technical documentation is marked up using SGML, XML, or some other markup language, the possibility exists of additional reconciliation of the different metadata schema. The well-defined EAD structure also makes possible the use of cross-walks to interface with other common metadata schema that might be relevant to the records (for example, geospatial metadata). Of course, the need to reconcile XML schema or map across different metadata seats is not unique to electronic records.

All this is not to say that EAD is ideal as it stands for describing electronic records. Several limitations need to be addressed in the next version of EAD if it is truly to accommodate electronic records:

⁹ J.E. Dryden, *Archival Description of Electronic Records: An Examination of Current Practices*, «Archivaria», 40 (1995), pp. 99-108.

Fig. 1
High-level model for the encoded archival description document type definition (Society of American Archivists Encoded Archival Description Working Group 1999)



1. EAD is strongest with regard to the description of the records once they are held in the archives. It is weak in how it supports records management, appraisal, and accessioning processes. More explicit attention needs to be paid to how records creation processes, records retention schedules, appraisal reports, accessioning procedures, and data quality reports are captured and tracked, as well as the relationships between the various agents associated with those processes. Moreover, it is likely that many archival records will never be physically accessioned by archives, but will remain in physical custody of the creators, with archivists setting and monitoring the requirements for appropriate control, as well as providing intellectual access to the records. EAD must be able to support non-custodial, as well as custodial programs for the archival management of electronic records.
2. There need to be more closely delineated data elements through which the procedural, temporal, structural, and semantic metadata accompanying the electronic records can be described, rather than consigning such materials to non-specific «bucket» elements such as <odd> and <add>. These elements and their values should be based upon lists of common types of documentation that accompany electronic records when they are created as well as when they are accessioned. The data elements also should have quality control attributes that indicate the extent to which the accuracy of each piece of documentation has been verified. It should also be possible to extract selected data elements from these metadata and place them in a relational table where they could be more easily be searched by the end user in addition to the finding aid.
3. Tracking and validating custodial history is integral to establishing the authenticity of records, and for electronic records this can be quite complex, especially if the archives takes over intellectual but not physical control of inactive records. The EAD <custodhist> (custodial history) element needs to be expanded to address this issue, in particular, non-custodial arrangements for archival electronic records.
4. Preservation processes and meticulous documentation of those processes are obviously critical not only for providing continued access to electronic records, but also for establishing and demonstrating the continued authenticity of those records (or of authentic copies of the records). Currently preservation information is bundled into a single EAD element <processinfo> (processing information), and as with <custodhist> this element needs to be expanded and further delineated to document preservation processes such as migration and emulation, the agents involved, and any effects that these might have upon the record.
5. Even with traditional records, many archivists find it difficult to make the necessary distinction between intellectual and physical levels of arrangement. Many electronic records can be arranged in multiple ways and, therefore, the concept of levels of arrangement may not be as relevant as possible arrangement schema. It needs to be possible through the EAD <arrangement> element

for users to identify the range of potential arrangements and views, and types of data extracts and redactions in order to be able to specify the one which they would like to use when accessing electronic records online or when ordering copies of them. This is a compelling reason to do more research into user needs so that any extensions to EAD are more user-driven.

6. As with museum objects, additional aspects of physical description may need to be incorporated into the <physdesc> element to allow for highly technical description of electronic media and formats. Some of these elements might correspond to those that were included in MARC MRDF.
7. For EAD in general, there is a need for a companion content standard and rules for developing authority files. Work on both of these aspects is currently underway in North America.

Conclusion

There is obviously much work to be done in the area of electronic records description, and EAD provides one important vehicle to do so. However, given the volume of electronic records already created and anticipated in future years, there must surely also be an increased emphasis on automating as many aspects of archival description as possible. This is where research and development projects such as those underway at the San Diego Supercomputer Center in partnership with the US National Archives and Records Administration are likely to make such an important contribution, as much because of their fundamental reconceptualizations of how archival information systems might function as because of the application of supercomputer computational capabilities.

One final caveat, however – almost all developments in archival description to date, even that of EAD, have occurred without systematic analysis of user needs and capabilities. As archival description, and even the complete archival record becomes increasingly available online to the general public without any archival reference mediation, it is going to be critical that we spend time examining the usefulness and usability of the materials we are providing to our users. Otherwise we may find that we have created a web of metadata and records that is so complex that it will have become impenetrable to most users.

1. EAD is strongest with regard to the description of the records once they are held in the archives. It is weak in how it supports records management, appraisal, and accessioning processes. More explicit attention needs to be paid to how records creation processes, records retention schedules, appraisal reports, accessioning procedures, and data quality reports are captured and tracked, as well as the relationships between the various agents associated with those processes. Moreover, it is likely that many archival records will never be physically accessioned by archives, but will remain in physical custody of the creators, with archivists setting and monitoring the requirements for appropriate control, as well as providing intellectual access to the records. EAD must be able to support non-custodial, as well as custodial programs for the archival management of electronic records.
2. There need to be more closely delineated data elements through which the procedural, temporal, structural, and semantic metadata accompanying the electronic records can be described, rather than consigning such materials to non-specific «bucket» elements such as <odd> and <add>. These elements and their values should be based upon lists of common types of documentation that accompany electronic records when they are created as well as when they are accessioned. The data elements also should have quality control attributes that indicate the extent to which the accuracy of each piece of documentation has been verified. It should also be possible to extract selected data elements from these metadata and place them in a relational table where they could be more easily be searched by the end user in addition to the finding aid.
3. Tracking and validating custodial history is integral to establishing the authenticity of records, and for electronic records this can be quite complex, especially if the archives takes over intellectual but not physical control of inactive records. The EAD <custodhist> (custodial history) element needs to be expanded to address this issue, in particular, non-custodial arrangements for archival electronic records.
4. Preservation processes and meticulous documentation of those processes are obviously critical not only for providing continued access to electronic records, but also for establishing and demonstrating the continued authenticity of those records (or of authentic copies of the records). Currently preservation information is bundled into a single EAD element <processinfo> (processing information), and as with <custodhist> this element needs to be expanded and further delineated to document preservation processes such as migration and emulation, the agents involved, and any effects that these might have upon the record.
5. Even with traditional records, many archivists find it difficult to make the necessary distinction between intellectual and physical levels of arrangement. Many electronic records can be arranged in multiple ways and, therefore, the concept of levels of arrangement may not be as relevant as possible arrangement schema. It needs to be possible through the EAD <arrangement> element

for users to identify the range of potential arrangements and views, and types of data extracts and redactions in order to be able to specify the one which they would like to use when accessing electronic records online or when ordering copies of them. This is a compelling reason to do more research into user needs so that any extensions to EAD are more user-driven.

6. As with museum objects, additional aspects of physical description may need to be incorporated into the <physdesc> element to allow for highly technical description of electronic media and formats. Some of these elements might correspond to those that were included in MARC MRDF.
7. For EAD in general, there is a need for a companion content standard and rules for developing authority files. Work on both of these aspects is currently underway in North America.

Conclusion

There is obviously much work to be done in the area of electronic records description, and EAD provides one important vehicle to do so. However, given the volume of electronic records already created and anticipated in future years, there must surely also be an increased emphasis on automating as many aspects of archival description as possible. This is where research and development projects such as those underway at the San Diego Supercomputer Center in partnership with the US National Archives and Records Administration are likely to make such an important contribution, as much because of their fundamental reconceptualizations of how archival information systems might function as because of the application of supercomputer computational capabilities.

One final caveat, however – almost all developments in archival description to date, even that of EAD, have occurred without systematic analysis of user needs and capabilities. As archival description, and even the complete archival record becomes increasingly available online to the general public without any archival reference mediation, it is going to be critical that we spend time examining the usefulness and usability of the materials we are providing to our users. Otherwise we may find that we have created a web of metadata and records that is so complex that it will have become impenetrable to most users.