European Commission

INSAR

INformation Summary on ARchives

EUROPEAN ARCHIVES NEWS

SUPPLEMENT IV (2000)

# Proceedings of the DLM-Forum on electronic records

## European citizens and electronic information: the memory of the Information Society

DLM-FORUM
Electronic Records

Brussels, 18-19 October 1999

- This publication contains the full texts in English and abstracts in German and French. These abstracts are placed directly after the full texts.
- The full texts for all contributions, in German, English and French, can be found on the DLM-Forum website: http://www.dlmforum.eu.org

---

- Diese Veröffentlichung enthält die vollständigen Texte in Englisch und Zusammenfassungen in Deutsch und Französisch. Die Zusammenfassungen befinden sich direkt hinter den vollständigen Texten.
- Die vollständigen Texte aller Beiträge sind in deutscher, englischer und französischer Sprache auf der DLM-Forum Website verfügbar: http://www.dlm-forum.eu.org

---

- Cette publication contient les textes complets en anglais et des résumés en allemand et en français. Les résumés se trouvent directement après les textes complets.
- Vous pouvez trouver les textes complets de toutes les contributions, en allemand, en anglais et en français, sur le site web du DLM-Forum: http://www.dlmforum.eu.org

EUROPEAN
ARCHIVES NEWS

# INSAR

- Secretariat-General of the European Commission

- Committee of preparation:
  Lino FACCO
  Hans HOFMANN
  Christina BECKERS
  Peter BERNINGER
  Jef SCHRAM
  Johanna MAAS

- For further information and a free copy of INSAR, please contact:
  European Commission
  Tel. (32-2) 29-56810
  (32-2) 29-56721
  Fax (32-2) 29-61095
  E-mail: archis@cec.eu.int
  Address: rue de la Loi 200
  Office: JECL 3/40
  B-1049 Brussels
  BELGIUM

# Proceedings

## of the DLM-Forum

## on electronic records

# European citizens and electronic information: the memory of the Information Society

Brussels, 18-19 October 1999

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server (http://europa.eu.int).

Cataloguing data can be found at the end of this publication.

Luxembourg: Office for Official Publications of the European Communities, 2000

ISBN 92-828-8806-1



*Printed in Belgium*

# Persistent object preservation: Advanced computing infrastructure for digital preservation ([1])

**Kenneth Thibodeau, Reagan Moore, Chaitanya Baru**

To preserve something is to maintain it intact and unchanged. We preserve records because we believe they have continuing value, with an expectation that this value will be realised in use. As the International Council on Archives' Committee on Electronic Records has stated, "The fundamental requirement for preservation is that the records remain physically intact, identified, and readable. 'Readable' electronic records are recoverable from storage for processing by a computer or presentation to humans."([2]) Preservation includes accessibility, but given the rapid obsolescence of information technology entails the probability that any digital object maintained unchanged for any length of time will become inaccessible.([3]) This tension between keeping electronic records unchanged and keeping them accessible necessitates identifying and adopting a preservation strategy which does not impact negatively the accessibility of the records.

The National Archives and Records Administration (NARA) of the United States is sponsoring research that articulates such a strategy and offers substantial promise for the development of a comprehensive archival system for electronic records. This research, cosponsored by the Advanced Research Projects Administration of the U.S. Department of Defense, and the U.S. Patent and Trademark Office, is being conducted by the National Partnership for Advanced Computational Infrastructure, led by San Diego Supercomputer Center (SDSC). The research is an extension of the Distributed Object Computation Testbed (DOCT), which provides a nation-wide infrastructure of supercomputing resources connected by high-speed networks supporting computation-intensive applications. This distributed infrastructure enables storage and management of massive data sets. The testbed supports preservation, access and dissemination requirements for large and diversified collections of compound, complex documents and information sets, including combinations of multiple file types and complex work units. What NARA has specifically added to DOCT is a focus on long-term preservation and access in accordance with archival standards, an extension of testbed technology to very diverse but ordered collections, and a requirement to demonstrate scalability of the solution.

## Empirical framework

The research project is motivated empirically by the continuing explosion in the quantity and the diversity of electronic records being created in the Government of the United States. For example, the Department of State will begin soon to transfer annually to NARA approximately one million diplomatic messages in electronic form. In 2001 more than twenty-five million e-mail messages are expected to be transferred from a single system in the Clinton White House. While NARA has increased its capacity to accession and preserve historically valuable electronic records by a factor of 500 over the past decade, existing systems cannot be expanded in an economically feasible manner to process the quantities of electronic records expected to be received in the near future.

The increasing diversity of forms and formats of electronic records constitutes an even greater challenge. The newer classes of computer applications, such as geographic information systems, object oriented data bases, dynamic World Wide Web pages, and virtual reality models, also tend to be increasingly complex. Proven methods for long-term preservation of electronic records are limited to a few, relatively simple forms. To preserve the great diversity of electronic records that are being created today and will be created in the future requires the development of methods applicable to their various formats.

## Organisational framework

NARA's motivation in sponsoring this research is clearly conditioned by its mission to preserve permanently valuable records of the United States Government. Accomplishing this mission requires solutions that adhere to the principles of archival science and embody the best practices of archival institutions. NARA needs to ensure that preservation methods for electronic records not only overcome the prob-

**Kenneth Thibodeau**

*Kenneth Thibodeau is Director of Electronic Records Programs at the National Archives and Records Administration (NARA). From 1988 until 1998, he was Director of NARA\'92s Center for Electronic Records. He oversees a number of initiatives aimed at achieving major progress in addressing the archival and records management challenges related to electronic records. In 1995, Ken was detailed from NARA to the Department of Defense, where he served as Director of the Records Management Task Force, which was set up to implement business process reengineering for records management. Prior to joining NARA in 1988, Ken was the first Chief of the Records Management Branch of the National Institutes of Health (NIH) with responsibilities for records management, office automation, privacy, and strategic planning for information resources management. Ken holds a bachelor's degree in history from Fordham University and earned a Ph.D. in the history and sociology of science from the University of Pennsylvania.*

(1) Kenneth Thibodeau is with the Modern Records Program of the National Archives and Records Administration, College Park, Maryland. Reagan Moore and Chaitanya Baru are with the San Diego Supercomputer Center, San Diego, California.
(2) International Council on Archives' Committee on Electronic Records. Guide for Managing Electronic Records from an Archival Perspective. 1997. P. 43
(3) It has been argued that the surest way of keeping electronic records unchanged and ensuring the possibility of retrieving them in their original forms is to maintain them in their original, native formats. However, original formats are almost always bound to the software, and sometimes to the hardware, used to create them. While there may be methods of perpetuating the original information technologies, such as emulation or simulation, application of such methods inherently will become more complex over time, and thus more risky. Moreover, success in applying these methods will tend to raise greater and greater barriers to using state of the art technology to find, retrieve, process, use, or communicate them. For example, many users would not be pleased if, in order to access digital objects that had been preserved across the last 30 years, they had to learn to use the PL1 programming language or Model 204 database software.

lems of obsolescence and media fragility, but also can be integrated into the performance of related archival functions including accessioning, description, and dissemination. NARA's requirement is to develop an information management architecture for preservation of electronic records that is scalable to millions of records; can accommodate a great variety of formats; and supports continuing access and authentic presentation of the preserved records into the indefinite future. NARA has needs performance and cost metrics for implementation of such an architecture.

### Technology framework

The research specifically addresses NARA's requirements, but the solutions it is developing derive from advances in computer science and technology that serve a wide variety of needs in government, scientific research, education, and other areas. NPACI is pursuing this research not as a special case, but in the context of a broad- scale effort in its Data Intensive Computing Environments thrust area, transferring advanced technology from independently funded, peer-reviewed computer science research activities. This collaborative program aims at the integration of distributed persistent digital archives, hierarchical storage systems, databases, data-handling systems, and digital libraries into integrated information repositories. The research is being applied across a variety of user communities including NASA, the California Digital Library, the Pacific Rim Digital Library, the Neuroscience Digital Library of Brain Images, the Art Museum Image Consortium, and many others. Technologies being applied in this research are seen as key elements in the next generation Internet, electronic commerce, and digital government.

### Conceptual framework

The conceptual framework for the archival research was articulated in Thibodeau's communication to the first DLM- Forum.([4]) The cornerstone of this framework is the recognition that electronic records are complex objects that inherit properties from several different classes of objects. All electronic records belong to two superclasses, that of digital objects on the one hand and of records on the other. Within the hierarchy of digital objects, all electronic records inherit certain properties from the subclass of physical files; that is, properties related to the media on which they are inscribed and the standards or conventions used in writing the files. As digital objects, electronic records also belong to one or more classes of software applications and inherit attributes and methods depending on whether they are word processing files, scanned document images, e-mail messages, multimedia files, etc. Within the superclass of record, electronic records inherit properties from their documentary form and from the archival fonds in which they belong. Each of these subclasses may be further subdivided into child classes and hierarchies. One advantage of an object-oriented approach is that it provides a coherent means of addressing methods as well as attributes.

There are two keys to implementing an object-oriented framework for preservation. First, we need to identify the specific classes to which each electronic record belongs; determine which of its attributes and methods are inherited from each of these classes; and evaluate which ones must be preserved. Segregating the essential from the inessential will enable us to concentrate resources on only those aspects of electronic records that are essential to their nature as records. Only essential aspects have to be preserved. For example, it is neither necessary nor possible to preserve the physical media on which electronic records are written as long as the records are copied intact to other media. Nonetheless, there are significant questions as to what other aspects of electronic records must be preserved. The international InterPARES research project is investigating, from the perspective of archival science, the requirements for preserving the authenticity of electronic records.([5])

**(4) K. Thibodeau. Boundaries and transformations: an object oriented strategy for the preservation of electronic records. Proccedings of the DLM-Forum on Electronic Records. 1996. Pp. 161-67.**
**(5) See http://www.InterPARES.org**

While the InterPARES research is in its early stages, we can assert that both the documentary nature of electronic records and their position within archival fonds need to be preserved across time. A record is a type of document. Clearly, we want to preserve the entire contents of a record and all essential aspects of its form or structure. In many cases, we may need to preserve the specific presentation of the record; that is, how it appears or sounds to humans. The presentation of electronic records is a method that is applied to the record, not an attribute captured in the stored record, as it is with traditional records stored on 'hard' media.

Preserving all of the essential attributes and methods of the document in itself would not be sufficient for archival preservation of records. What differentiates records from other forms of documentary materials is that a record has a specified provenance and an archival bond which situates the record in relation to the other records of the same creator. The archival fonds is the immediate and enduring context of a record; therefore, we must preserve not only the documents themselves, but also the relationships among documents within an archival fonds. With traditional records, the relationship of a record to other records in an archival fonds is established, and can be maintained, by the physical placement of the record in a collection, supplemented by archival description of the fonds. But the archival bond of an electronic record may be defined only logically and realised only by software. Therefore, an archival system for preserving electronic records must not only be capable of capturing and sustaining the contents, forms and methods of individual records, but also it must be able to capture and instantiate the structures of collections of records. The archival bond may be captured in metadata which specifies the relationship between an electronic record and other records in the archival fonds, but the instantiation of an archival series or fonds as an ordered collection entails the application of one or more methods.

The second key to an object-oriented approach is that solutions should be sought and applied at the highest applicable level of the class hierarchies. For example, if we consider the documentary form of e-mail, all e-mail messages contain data identifying a sender, one or more recipients, a subject, and a date of transmission. These elements must be preserved as part of an e-mail record. Applying the "highest level" rule, it would be sufficient to preserve the identities of the sender and recipient(s) as such, and to distinguish them from names that simply appear in the body of the message. Absent any other applicable criteria, it would not be essential to preserve the methods that the original e-mail software used to identify sender and recipients; therefore, the same method could be used to preserve these elements in all e-mail records regardless of the original application software. Building and applying solutions at the highest level of the class hierarchy increases efficiency and improves management of the risks involved in preserving electronic records.

## Collection-based long-term preservation

The architecture that SDSC is developing includes infrastructure, language, data flow, data control, and implementation at levels ranging from physical storage through presentation for human reading and interpretation. The architecture is consistent with the ISO Open Archival Information System Reference Model (OAIS).([6]) SDSC has described its technical strategy as "Collection-based Long-term Preservation." There are four basic aspects of this strategy which merit attention: first, it is collection based; second, it covers archival processes from transfer of records to the archives through long-term access; third, it provides a consistent and comprehensive approach to a wide variety of electronic records; and fourth, it is infrastructure independent.

### Collection basis

Archives preserve collections of records respecting both provenance and original order. The collection-based, object oriented preservation strategy is perfectly suited to the archival requirements for preserving and providing access to heterogeneous collections of electronic records. This strategy preserves digital records as members of collections. The system preserves both individual electronic records and the information required to assemble and arrange the records in their original order. It uses this information dynamically to recreate the original collections when records are retrieved. It can accommodate both the hierarchical levels of arrangement that are typical of traditional archives and the complex structures that may apply to collections of electronic records, such as a database or geographic information system.

### Archival processing

The object oriented preservation strategy enables us to envision an electronic records archives system supporting archival processing for the accessioning or ingest of records into the archives, long-term preservation, information discovery against the preserved collections; and the presentation of the retrieved objects. From a user perspective, the system comprises two principal types of components: structural elements and tool sets. The structural components provide capability and capacity for performing the core archival functions of ingest, preservation, and access. They are virtually places where archival work may be done. The components for ingest and access are called the accessioning workbench and the reference workbench. Records are maintained over time in a component called the repository. Tool sets will be specialised applications optimised for specific tasks or processes and used on the accessioning and reference workbenches. ([7])

**(6) Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-W-4.0. September 17, 1998.**
**(7) In another project, at the Georgia Tech Research Institute, NARA has sponsored research on technologies applicable in building tool sets for archival processing of electronic records in cases where the records had not been subject to proper records management, and also for screening and redaction of records that may contain sensitive information. William E. Underwood. Analysis of Presidential Electronic Records: Final Report to the National Archives and Records Administration. September 1999.**

In the ingest process, the transferred collection is decomposed into the individual records or data objects which it contains, and the context and structure of each object are characterised by metadata. In the prototypes, a Document Type Definition (DTD) in eXtensible Markup Language (XML) is identified or, if necessary, derived for each class of object defined by documentary form; the DTD is applied and each object is annotated with the appropriate metadata using standard XML markup. This methods is applied to objects including individual documents, parts of documents such as tables or graphics included in textual documents, and collections of records. As with individual records, a DTD is constructed to capture the structure of a collection. Once encapsulated through application of a DTD, a collection becomes a discrete object which can itself be a member of a collection, thus supporting hierarchical arrangement. The collection-level DTD represents the hierarchical structure of the collection. In addition, a Data Definition Language (DDL) is used to describe the logical relationship between attributes of the collection and is mapped to the DTD. The research is investigating the characterisation of presentation interfaces as semi-structured data organised through the use of eXtensible Style Language (XSL), or other similar technology.([8]) The integrity of the ingest process can be, and has been, demonstrated by reversing it; that is, producing from the XML representation a copy in the original format and then comparing this output copy with the original. The ingest process produces metadata encapsulated records in a standardised format for archival storage and for subsequent retrieval and use.

The archival storage phase of the process includes not only deposit and maintenance of the collections of records in the repository, but also the capture and maintenance of the archival metadata needed to guarantee the integrity of the preserved records and to enable subsequent retrieval. The repository includes:

- the electronic records encapsulated within DTDs,
- the DTD for each genre of records in the archives,
- the DTD for each collection of records in the archives,
- the DDL for each collection,
- the mapping between each collection-level DTD and the corresponding DDL, and
- XSL style sheets defining how each object is to be presented.

An object may be stored in its original, native format, with a simple DTD wrapper identifying the object and its format. This provides an option for preserving objects in two formats (native format with external wrapper, or well-formed XML object) to offset any uncertainty in representation of the schemas of documents and collections via XML and DDL. In either case, the stored objects are aggregated in large containers, currently physical files on the order of 150 megabytes. This aggregation minimises the number of physical entities that must be managed in the repository and simplifies storage management because all containers have identical wrappers. When storage media or storage systems need to be replaced, the containers can be moved to the new media or system in a highly uniform process without regard to the objects stored in the containers.

Access to preserved records is achieved through dynamic reconstruction of the relevant collection of records, dynamic generation of a user interface to support queries against that collection, and dynamic generation of the code required to execute queries. The system rebuilds collections by retrieving the metadata for the target collection and dynamically creating a database, currently using object-relational database technology, to implement the schemas applicable to the collection. The data objects comprising the collection are retrieved from storage and loaded into this database through procedures that convert the XML representation of the objects and collections into a database representation. Standard Query Language is used to retrieve records from the collection database. The system can generate the required SQL statements dynamically by making use of the explicit definition of the structure of a collection contained in the metadata, including the organisation of attributes, their type definitions and their sizes.

**(8) San Diego Supercomputer Center. Persistent Archives. http://www.sdsc.edu/NARA/Publications/collections.html**

Once retrieved, presentation of the records is accomplished by dynamically creating the presentation interface for each record or collection through the application of the appropriate XSL style sheets. The presentation interface supports retrieval of original records and of the metadata about each record or collection of records. The process of dynamically rebuilding the data collection from the individual data objects stored in the repository and dynamically constructing the presentation interface for the digital objects discovered through a query effectively demonstrates the integrity of the system for preserving and providing access to authentic electronic records.

While the project is in its early stages, research to date has included prototype demonstrations of all of the archival processes described against a heterogeneous test corpus. For example, a collection of one million e-mail messages was ingested, stored and dynamically rebuilt within one day. The automated steps in this prototype included assembling the collection, tagging each message using XML, refining the DTD as new attributes were discovered in the collection, storing the objects within the archival repository, retrieving the stored objects and rearranging them as a collection, indexing the restored collection, presenting the collection through a Web interface, and supporting queries against it.([9])

**Consistent, comprehensive approach**

In the object-oriented strategy, both individual records and collections of records are managed as digital objects through metadata. The strategy has been applied in prototypes using a test corpus that reflects the diversity of electronic records confronting archival institutions. The test corpus includes:

- databases representing a 25 year span of database technology,
- heterogeneous collections of files created in office automation environments, including the U.S. Congress and the Department of Defense Records Management Application,
- a geographic information system comprising almost 50.000 files,
- 1.000.000 e-mail messages,
- 2.000.000 electronic patent case files, and
- 10.000 digital images of art works.

The applicability of XML DTDs and XSL style sheets to the variety of genres of electronic records in the test corpus produces a collection of digital objects which is very homogeneous for purposes of repository management and maintenance. This greatly simplifies the long-term preservation of electronic records. The components of the system whose function is to transport the preserved records over time do not need to be concerned with the meaning of the contents, the internal form or formats of the records, their position in an archival fonds, or their presentation to researchers. The preservation subsystem simply needs to preserve the stored objects and their related metadata as they were originally deposited in the archives.

Persistence is achieved through identification of metadata for all attributes related to digital object properties and collection organisation. During the ingest process, metadata is automatically generated and captured for each genre of records and each collection of records. There is great flexibility in the types of metadata that can be captured. Metadata can also be captured about individual records, for example, user annotations. Beyond the structural metadata typically expressed in DTDs, metadata about provenance and access can be captured.

The metadata is managed using a metadata catalog, currently the Extensible Metadata Catalog (EMCAT) which SDSC developed at earlier stage of the DOCT project. The metadata catalog is itself a database comprising three different levels of metadata:

> Digital object metadata about type, formats, provenance, ingestion protocols, usage methods, and domain-specific metadata and related data, and information discovery and presentation information;
>
> System-level metadata about storage system characteristics, authentication and encryption, access and audit control, replication, partitioning and containers of data sets, and location transparencies; and
>
> Schema-level metadata defining schema attributes, ontology, indexing and relationships among attributes, evolution mechanisms for data and schema, and federation and integration information.

Access to preserved collections of records is provided through the intermediary of the Storage Request Broker (SRB), another software component developed by SDSC. The SRB is middleware that provides distributed clients with uniform access to the preserved collections. These collections may be stored in distributed and diverse storage systems in a variety of computing platforms. By providing the client with a logical view of the preserved records, SRB eliminates the need for the client to deal with any of the specifics of the operating environments or storage systems, including

(9) Reagan Moore, Chaitan Baru, Amarnath Gupta, Bertram Ludaescher, Richard Marciano, Arcot Rajasekar. Collection-Based Long-Term Preservation. San Diego, California. San Diego Supercomputer Center. June 1999.

where the records are physically located. The SRB also provides authentication, encryption, access control and auditing capabilities. SRB uses the meta-information contained in EMCAT to access stored collections.(10)

### Infrastructure independence

The object oriented preservation strategy addresses technological obsolescence by making the archives essentially independent of the technological infrastructure on which it resides. Infrastructure independence is achieved with respect to both the preserved collections of records and the specific hardware and software components used to implement the archives. The records and collections are preserved as representations in the form of XML encapsulated objects. XML DTDs, DDLs and XSL style sheets enable the dynamic reconstruction and presentation of collections of records on target technologies which need not be similar to, or congruent with, the technology originally used to create the records or the collections. f the XML standard became obsolete, it would not be necessary to migrate the collections stored in the repository to a new. It would only be necessary to develop a new interface that could translate the XML metadata to a form acceptable to the target technology.

With respect to the technology, the architecture includes a migration strategy for upgrading any infrastructure component, including new media, new storage software, new collection management databases and new presentation displays. The metadata catalog is critical to the migration strategy. System-level metadata provides location transparency, access transparency and protocol transparency. Schema-level metadata provides a way to migrate collections to new technology and to federate data collections.

Infrastructure independence is specifically embodied in the emergent design of a system for electronic records archives consisting of structural components and tool sets. This is readily apparent in the concept of a tool set as an application which will ideally "snap in" to the workbench where it is used. Tool sets can be developed for any required processing provided they can communicate with the workbench's standard interface. A similar situation exists in SDSC's design of the three structural components. The accessioning workbench, the repository, and the reference workbench can be modified independently as long as interacting components can address a standard application programming interface.

Infrastructure independence has been demonstrated at the San Diego Supercomputer Center. SDSC replaces its storage subsystem, including media and devices, every three years as a cost-saving method. Because of improved efficiency and reduced costs of new technology, it is actually less expensive to move all of their holdings to new technology than it would be simply to leave the old storage system in place. When the storage system is replaced , the holdings are simply moved from the old system to the new, without reformatting the digital objects stored in the system. This practice belies the common perception that obsolescence is a complicating factor adding to the costs of digital preservation.

Infrastructure independence offers additional potential benefits. Archival institutions have a basic responsibility for preserving and presenting authentic records, but they also serve customers with varying needs and interests. The object oriented preservation strategy provides the capability for meeting requirements for authenticity, but it also makes it possible to construct user interfaces and information delivery systems that are tailored to the needs of specific clienteles. As an example, many government archives around the world find that their users are predominantly people researching family history. Without changing the architecture of the electronic records repository or the reference workbench, it would be possible to construct and deploy an access tool set optimised for genealogical research.

**(10) C. Baru et al. A Data handling Architecture for a Prototype Federal Application. Proceedings of the IEEE Conference on Mass Storage Systems. College Park, MD. March 1998.**

## Conclusion

This paper provides only a very summary view into the electronic records archives research project. This research is still in progress and many questions need to be addressed in order to move from research and development to implementation. But substantial progress has been made, sufficient to justify optimism that it is possible technically and economically to develop an information management architecture that can satisfy archival requirements for maintaining electronic records both intact and accessible.

# Persistent Object Preservation: Fortgeschrittene DV-Infrastruktur für die digitale Konservierung

**Kenneth Thibodeau, Reagan Moore und Chaitanya Baru**

Auf dem ersten DLM-Forum legte Thibodeau einen Vorschlag für eine objektorientierte Strategie zur Konservierung von elektronischen Aufzeichnungen vor (K. Thibodeau. Grenzen und Transformationen – eine objektorientierte Strategie für die Konservierung elektronischen Archivguts, Vorträge und Ergebnisse des DLM-Forums, S. 161-167.) Wir wollen auf dem zweiten DLM-Forum beschreiben, wie diese Strategie bei Forschungs- und Entwicklungsarbeiten zur Anwendung fortgeschrittener EDV-Möglichkeiten auf die archivische Konservierung elektronischer Unterlagen verfolgt wird. Diese Arbeiten laufen im Rahmen einer Kooperation zwischen der National Archives and Records Administration, der Defense Advanced Research Projects Administration und dem Patent and Trademarks Office der USA und werden durch das San Diego Supercomputer Center durchgeführt.

Der Beitrag wird die Motive des Projekts, den Forschungszweck und die erwarteten Ergebnisse beschreiben. Er wird über den Stand der Festlegung einer Informationsmanagement-Architektur für Dauerarchive elektronischer Aufzeichnungen sowie über den Stand der Ausweisung einer Leistungs- und Kostenmatrix für die Umsetzung einer derartigen Architektur berichten. Ebenfalls beschrieben werden Praxistests zur Bestätigung der Nutzbarkeit des technischen Ansatzes durch Aufnahme verschiedener sehr großer (mehrere Millionen Aufzeichnungen) und sehr unterschiedlicher Sammlungen digitaler Objekte sowie durch die Übertragung und dynamische Neuerstellung von Sammlungen über technologische Grenzen hinweg.

Außerdem wird in dem Beitrag analysiert, wie bei diesem technologischen Ansatz grundlegende archivische Anforderungen wie Achtung der Provenienz und der ursprünglichen Ordnung von Aufzeichnungen sowie Aufbewahrung und Retrieval authentischer Einzelakten Berücksichtigung finden.

# Conservation permanente des objets : infrastructure informatique avancée pour la conservation des DLM

**Kenneth Thibodeau, Reagan Moore et Chaitanya Baru**

Lors du premier DLM-Forum, nous avions présenté une proposition de stratégie orientée objet pour conserver les documents électroniques (K. Thibodeau : "Limites et transformations : une stratégie orientée objet pour la conservation des documents électroniques". Actes du DLM-Forum, pp. 161-167). Lors de ce deuxième forum, nous nous proposons de montrer la manière dont des initiatives de recherche et de développement, visant à appliquer des capacités informatiques de pointe à la conservation de documents électroniques dans des archives, poursuivent cette stratégie. Ces travaux de recherche sont menés en collaboration entre la NARA (National Archives and Records Administration), la Defense Advanced Research Projects Administration et l'Office des brevets américain (PTO). Le Supercomputer Center de San Diego en assure la direction.

L'exposé présentera les motivations du projet, la finalité des travaux de recherche et les résultats escomptés. Il rendra compte des progrès réalisés dans la définition d'une architecture de gestion de l'information pour l'archivage permanent de DLM et dans la formulation d'un système de mesure des performances et du coût de mise en œuvre d'une telle architecture. L'exposé présentera également des tests en conditions réelles démontrant la validité de l'approche technique suivie, par l'absorption de divers fonds et collections très fournis (comportant plusieurs millions de dossiers) d'objets numériques et par le transfert et la reconstitution dynamique de collections à travers les barrières technologiques.

L'exposé analysera également comment cette approche technologique satisfait à des critères archivistiques fondamentaux, tels que le respect de la provenance des fonds et l'ordre des dossiers ainsi que la préservation et l'accès à des documents authentiques.