Taking Control:
Managing the Records Lifecycle
in an Automated Environment

Briefing for NARA Staff

Ken Thibodeau

March 19, 2003

NARA Needs to Manage Records

- Transferred to its physical custody
- Accepted in its legal custody
- Lifecycle management includes
  - Managing the records
  - Managing sets of records (files, series, etc.)
  - Managing the transactions in which records are involved.

Records

- All records are documents (= units of recorded information)

- Not all documents are records

- The difference between a document and a record is that the relationship between a record and its creator and the activity in which it was created is specified.

Electronic Records

- An electronic record is a document encoded in digital form, requiring a computer for processing

- There is no necessary or fixed relationship between or among the record, document or digital properties of an electronic record.

ERA
Electronic Records Archives

Dimensions of an Electronic Record

• **Record**
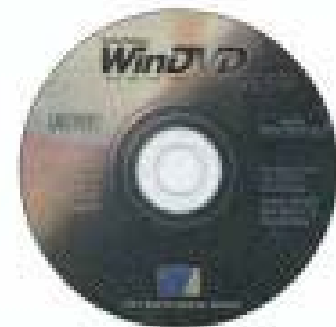
• Provenance

• Relationship to Creator

• Relationship to Creating Activity

• Relationship to Other Records

• Arrangement

• Archival Bond

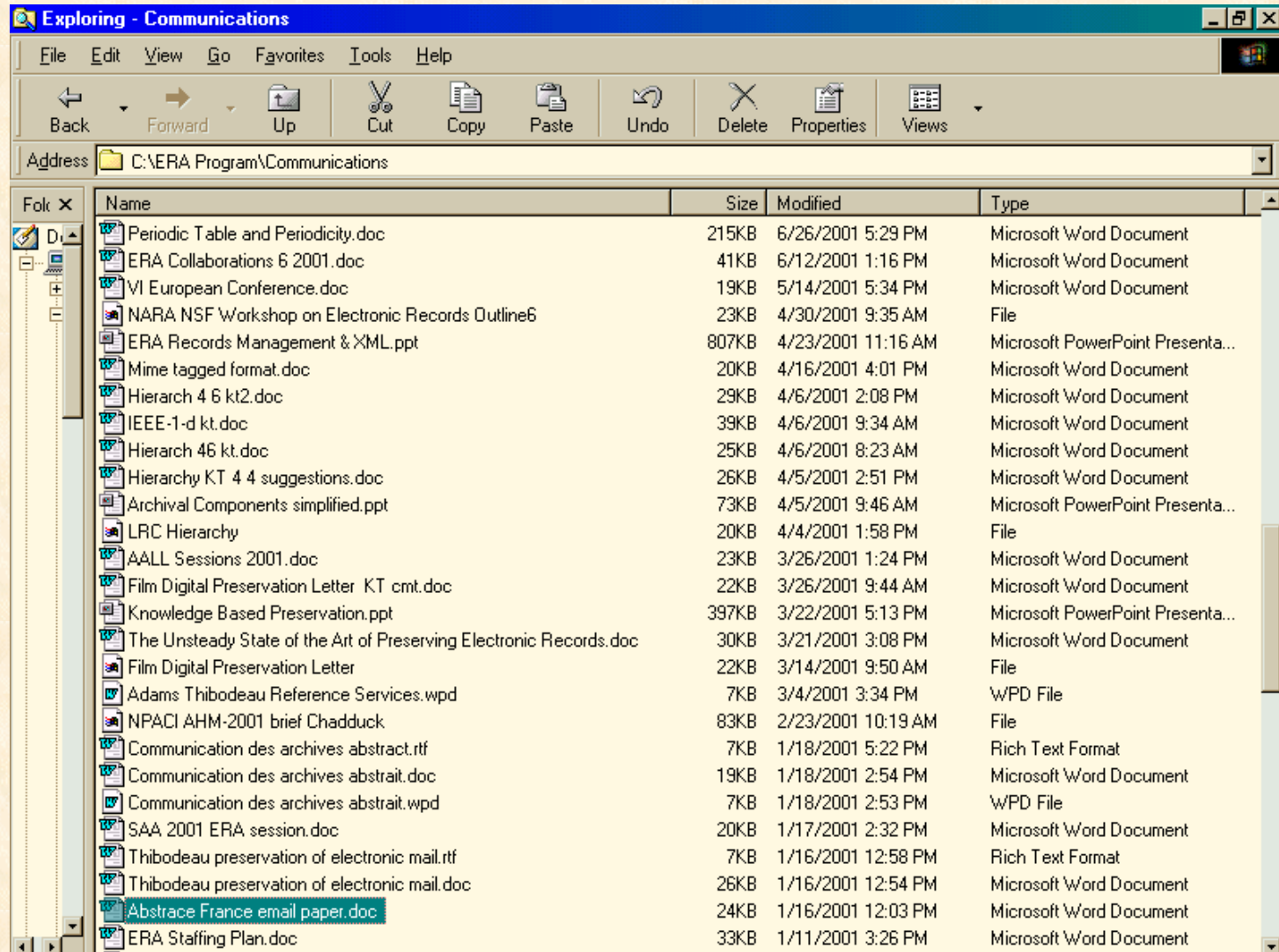•**Document**

• Content

• Structure

• Presentation

•**Digital Object**

• Representation or encoding of data in binary form

• Logical Structure of encoded data

• Physical Inscription of the encoding on a medium of storage or transmission

Inscription on Physical media

Inscription in a Physical File

Physical inscription of a Document

- word processing file
- word processing file that contains a spreadsheet file
- word processing file containing a pointer to
  - a spreadsheet file, or
  - a digital picture of the table

Digital Encoding of a Document

AGENDA

Date	Time	Agenda Items
Thursday, December 5th	9:30-12:30	• Welcome • Advisory Board introductions • Agenda summary
		• Discussion of the project's intellectual framework, including ○ Deadlines and level of detail for integrated cross-domain, cross-focus models for electronic records creation, management, appraisal, and long-term preservation and access
	12:30-14:00	Lunch
	14:00-17:00	• Continuation of intellectual framework discussion
Friday, December 6th	9:30-12:30	• Discussion of the latest version of the milestones document and its implications, relating it to the framework
	12:30-14:00	Lunch
	14:00-17:00	• Discussion of InterPARES 2 investing in prototyping of proposed preservation methods from sources outside of the project • Review of modeling software analysis and report • Discussion of data models
Saturday, December 7th	9:30-12:30	• Discussion and possible approval of new case study proposals • Review of ongoing case studies
	12:30-14:00	Lunch
	14:00-17:00	• Discussion of the role of the Advisory Board and discussion of its comments • Acceptance of new members and review of team members • Administrative issues

- word processing file
- Adobe ".pdf" file
- spreadsheet
- database report
- HTML web page
- scanned image

Logical Structure of a Document

Figure 3: Illustration of a "Table Sorting" behavior. A table is highlighted in the image, and sorted by clicking on the "TODAY" header. Note that highlighting of components (in the case, of matched search term regions) is preserved as the image is manipulated.

4 The Multivalent Document Architecture

We present a brief overview of the multivalent document architecture. A more detailed description can be found in [Phe98].

With layers and behaviors of arbitrary type coming together from multiple sources, a key problem is their coherent composition into a single conceptual document for the user. This integration is accomplished in the multivalent document architecture by several features:

1. A well-defined suite of protocols (implemented as method signatures in Java) to which behaviors should conform. The model's built-in logic promises to compose conforming behaviors.

2. A separation of structural document content from media-dependent elements.

- Logical Structure is the way the computer organizes the data that comprises the document.
  - Text
    - alphanumeric characters, paragraphs, headings, pages, etc.
  - Spreadsheet
    - pages, rows, columns, cells, characters, formulas
  - Image
    - Rectangle of black, white & color points (pixels)

Logical v. Conceptual Structure

- Logical Structure is the way the computer organizes the data that comprises the document.

- Conceptual Structure is the the organization of a document as perceived by a person

Electronic Records

- An electronic record is a document encoded in digital form, requiring a computer for processing

- There is no necessary or fixed relationship between or among the record, document or digital properties of an electronic record.

- **It is possible, and may be desirable or necessary, to vary digital properties in order to preserve electronic records.**

To Manage Electronic Records Efficiently and Effectively

- Automate processing

- Incorporate laws, regulations, policies, and sound archival and records management principles in the system.

- Make electronic records *self-describing* and *self-validating* to facilitate processing and management

Self-presenting documents

Plain text view of self presenting Document

```
<body bgcolor="#FFFFFF" text="#000000" link="#0000FF" vlink="#990000" alink="#666699" leftmargin=
<table width="723" border="0" cellpadding="0" cellspacing="0" height="1112">
  <tr>
    <td valign="top" colspan="3" height="2">
      <table width="744" border="0" cellpadding="0" cellspacing="0" height="152" bgcolor="#003399
        <tr>
          <td width="113" height="108" valign="middle" align="center" bgcolor="#003399">
            <div align="right"><a href="/omb/index.html"><img src="/omb/images/seal_omb_lines_blu
          </td>
          <td width="520" rowspan="2" valign="middle" align="center" bgcolor="#003399">
            <div align="center"><img src="/omb/images/logo_omb.gif" width="512" height="128" alt=
          </td>
          <td valign="middle" width="113" align="center" bgcolor="#003399" height="108">
            <div align="left"><a href="http://www.whitehouse.gov/index.html"><img src="/omb/image
          </td>
          <td width="1"></td>
        </tr>
        <tr>
          <td height="29" valign="middle" align="center" bgcolor="#003399" width="113">
            <div align="center"><font size="1" color="#FFFFFF" face="Verdana, Arial, Helvetica, s
          </td>
          <td valign="middle" align="center" bgcolor="#003399" width="113">
            <div align="center"><font size="1" color="#FFFFFF" face="Verdana, Arial, Helvetica, s
          </td>
          <td width="1"></td>
        </tr>
        <tr valign="middle">
          <td colspan="3" valign="middle" align="center">
            <table width="744" border="0" cellspacing="1" cellpadding="4" height="8" align="cente
              <tr bgcolor="#ff0000">
                <td height="8">
                  <table width="737" border="0" cellspacing="0" cellpadding="0" height="15">
                    <tr bgcolor="#FFFFFF">
                      <form name="seek" method="GET" action="/omb/query.html">
                        <td width="217" height="12" align="left" valign="bottom">
```

ERA
Electronic Records Archives

Self-describing Electronic Records

Document Type: Transmittal Memorandum

February 10, 1998

Document Date →

Directive

CIRCULAR NO. A-119

Revised

Addressees

Status

MEMORANDUM FOR HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

SUBJECT: Federal Participation in the Development and Use of Voluntary Consensus Standards and in Conformity Assessment Activities

Subject

Revised OMB Circular A-119 establishes policies on Federal use and development of voluntary consensus standards and on conformity assessment activities. Pub. L. 104-113, the "National Technology Transfer and Advancement Act of 1995," codified existing policies in A-119, established reporting requirements, and authorized the National Institute of Standards and Technology to coordinate conformity assessment activities of the agencies. OMB is issuing this revision of the Circular in order to make the terminology of the Circular consistent with the National Technology Transfer and Advancement Act of 1995, to issue guidance to the agencies on making their reports to OMB, to direct the Secretary of Commerce to issue policy guidance for conformity assessment, and to make changes for clarity.

Legal Basis

Message Body

Author → Franklin D. Raines

Action Officers

ERA
Electronic Records Archives

Sample Self-Describing Document

<Document type: transmittal memorandum>

<Document Date:> February 10, 1998</>
<Directive Transmitted:> Circular A-119</>
<Addressees:>Heads Of Executive Departments And Agencies</> **<Subject:>**Federal Participation in the Development and Use of Voluntary Consensus Standards and Conformity Activities </> < **Message Body:>**Revised OMB Circular A-119 establishes policies on federal use

ERA
Electronic Records Archives

Self-validating Electronic Records

- A **self-describing document** has meaningful indicators of what the document is and/or contains

  - Ideally, the indicators ("tags") are simple and clear enough to be correctly interpreted by any computer or by a person

- When the tags can be used to compare a self-describing document to a model ("template") of what it should be, it can be a **self-validating document**

Template Example

Document Type: Transmittal Memorandum

Document Date

February 10, 1998

Directive

CIRCULAR NO. A-119

Revised

Status

Addressees

MEMORANDUM FOR HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

SUBJECT: Federal Participation in the Development and Use of Voluntary Consensus Standards and in Conformity Assessment Activities

Subject

Revised OMB Circular A-119 establishes policies on Federal use and development of voluntary consensus standards and on conformity assessment activities. Pub. L. 104-113, the "National Technology Transfer and Advancement Act of 1995," codified existing policies in A-119, established reporting requirements, and authorized the National Institute of Standards and Technology to coordinate conformity assessment activities of the agencies. OMB is issuing this revision of the Circular in order to make the terminology of the Circular consistent with the National Technology Transfer and Advancement Act of 1995, to issue guidance to the agencies on making their reports to OMB, to direct the Secretary of Commerce to issue policy guidance for conformity assessment, and to make changes for clarity.

Message Body

Author → Franklin D. Raines

- A "Transmittal Memorandum" must contain
  - Reference to the directive which it transmits
  - A subject header
  - Indication of addressees
  - Identity of the transmitting official
  - Date transmitted by the official
  - A message body
- The message body may identify the legal basis and one or more action officers

If the template is in executable form, the computer can determine automatically if a record is what it should be.

Controls on Automated Lifecycle Management

- Methods to ensure that
  - What must or must not happen in any transaction or process does or does not happen, and
  - What must be true about any record or sets of records remains true.
- There can be different types of controls; e.g.,
  - Workflow Management
  - Business Rules
  - Templates
  - Preservation Strategies

Automated Controls: Business Rules

- "Records should not be transferred to the National Archives unless they have been appraised as permanent in an authorized records schedule."

- "The records that NARA preserves must be the same records, in all essential respects, as those transferred to it; i.e., NARA must preserve authentic records."

  - "State Department diplomatic messages have the documentary form of a telegram."

ERA
Electronic Records Archives

Automated Controls: "Template"

- An abstraction or articulation of the properties of a record, set of records, or a transaction involving records, which must be controlled.

  - The template for all records requires that the provenance, date, and archival bond be specified.

  - The template for State Dept. diplomatic messages requires that all content be plain text.

  - The template for all transfers of records to the National Archives is the S.F. 258

Hierarchy of Controls

Tier I: NARA STANDARDS

Apply to Abstract Classes of Records and Sets of
Records and to Lifecycle Transactions

Create: Preserver

Register: Preserver

↕ Conform to NARA standards

Tier II: Top Level Producer Templates
Apply to either abstract or real Classes &
Sets of Records

Create: Record Producer

Register: Appraiser

↕ Conform to Tier II templates

Tier III: Subsidiary Producer Templates
Apply to real Classes of Records and Sets of Records

Create: Record Producer

Register: Record Producer

Guidelines

Specifics

ERA
Electronic Records Archives

Records Template Hierarchy

Files Template Hierarchy

NARA STANDARDS

File

Tier I
Subject File | Case File | xxx File

Tier II
INS Case Files | NIH Grant Files | NIH Medical Records | DoD OMPF

Tier III
Visa Holders | Naturalization | Resident Alien | Project Grant | Center Grant | Army OMPFs | ○○○ | Navy OMPFs

Naturalization:
- 1975-2003
- 2004 -

ERA
Electronic Records Archives

Top "FILE" Standard

- How is a file identified?
- Where is the file located in the filing system?
- What does the file cover?
  - E.g., a topic, a case
- What types of records must, may, or may not be included in the file?
- What are the opening and closing dates of the file?
- What is the file's disposition?

ERA
Electronic Records Archives

Case File Standard

File → Case File

- File
  - Identifier
  - Location
  - Coverage
  - Required Records

  - Dates

  - Disposition Authority

- Required Elements
  - Case Identifier
  - Subseries of
  - Nature of case
  - Case Opening Record
  - Case Closing Record
  - Case Start Date
  - Case End Date
  - Disposition Authority

- Optional Elements
  - Producer Required Records
  - Producer Optional Records
  - Producer Prohibited Records
  - Cross Reference
  - Unspecified Records
  - ....

ERA
Electronic Records Archives

Control

Transaction

Process

Time

ERA System

Records Schedule

File & Records Templates

Transfer Controls

What should be transferred?

When should the transfer happen?

In what form will the records be transferred?

Who authorizes the transfer?

Agency Records

Transferred Records

Transfer of Records

Records Center
Service Agreement

Terms & Conditions

Records
in FRC

Transferred
Records

Transfer of Records

Records
Schedule

~~File &
Records
Template~~

Records Center
Service Agreement

Transfer records on CD-ROMs 1
year after close out

Destroy after 6 years

Return boxes or CDs on demand

Box lists

CD labeling
convention

Records
in FRC

Center Services

Records Schedule

Records Center Service Agreement

Transfer records on CD-ROMs 1 year after close out

Destroy after 6 years

Return boxes or CDs, or FTP individual files on demand

Box lists

CD labeling convention

File directories

Records in FRC

Center Services

Accession of Records

Case File: d-OMPF Example

F i l e

R e c o r d

NARA Standard	DoD Template	Instance
Case File	Official Military Personnel File	**Header:** SSN=xxx-xx-xxxx
File Identifier	SSN	Doc Class=DD 214
….	….	Record Date=*dd/mm/yyyy*
Closing Record	Certificate of Release or Discharge from Active Duty	….
Document Class: Form		
Form Number	DD 214	
Form Owner	DoD	
Form Version	nn	
Version Issue Date	198x	
Version End Date	199x	
	Form specification	
Digital Component		
	Header label	
Component Name	ASCII	
Data type	Page image	
	TIF IV	

ERA
Electronic Records Archives

Case File Data:

File ID = SSN=xxx-xx-xxxx

....

Closing Records = Doc Class=DD 214

Closing Record Date = *dd/mm/yyyy*

....

Image Header: SSN=xxx-xx-xxxx

Doc Class=DD 214

Record Date=*dd/mm/yyyy*

....

DoD DIHRMS Personnel Data Model

Virtual DD 214 Database Logical Model

Retrieve DD 214 from Database

Use logical model to materialize Virtual DD 214 Database

DOMPF Series Template

DD 214 Form Template

Retrieve Requested Person's Record from Storage

DOMPF Case File Template File Identifier

Place Retrieved Data in DD 214 Database

Apply DD 214 Form to data in Database

DD 214 Form Style Sheet

Output DD 214

Automated Controls: Preservation Strategies

- Methods applied to specified sets of records to achieve NARA's overall goal of preserving electronic records free from dependence on specific hardware or software

  – "Character based text files will be preserved in XML format, with associated eXtensible Style Sheets used to preserve appearance"

  – "Character based text files will be preserved in XML format, with an associated image file to preserve appearance"

ERA
Electronic Records Archives

Preservation Strategy

- A plan for ensuring the continuing authenticity and accessibility of records for as long as they need to be preserved
    - Applies to an archival aggregate of records
    - Preserves the essential properties of the records and the aggregates of records
    - Takes into account the digital properties and how they relate to record properties.
- May be determined hierarchically

Hypothetical Preservation Strategies:
Case File Standard

- If *documents* are received as *separate files*, transfer must include *data* identifying all *records* in the file, the order of each *record*, and the start and end dates of the *case file*.

- If *documents* are stored within a database, the logical model of the database must identify how a case file, all records that belong in a case file, and their order within the file are mapped to the logical database model.

- …..

■ Record Object	■ Document	■ Digital Component

Hypothetical Preservation Strategies:
Case Opening Record Standard

- If *document* is stored as *digital image*, *metadata* must indicate record name, date, and status as opening record.

- If *document* is stored as *character based file*, record name, date, and status as opening record must be tagged in the record.

- …..

■ Record Object	■ Document	■ Digital Component

ERA
Electronic Records Archives

Hypothetical Preservation Strategies:
High Reference Use Standard

- If *textual documents* are **not** stored as *self-describing documents,* convert to character encoding and tag significant elements of content (e.g. author, title, date, document summary, filing code) for preservation.

- Transform from preservation format to target format that can be readily accessed online by researchers.

- …..

 Record Object   Document   Digital Component

Controls on Automated Lifecycle Management

- Must be implemented in ERA
  - to manage records
  - to manage ordered aggregates of records
  - to manage transactions involving records

Controls on Records

- Define and identify the essential characteristics of each record

  – Content, structure, appearance, provenance, position in original order

- Determine Preservation Strategies

- Enhance possibilities for search & retrieval

- Facilitate review, redaction, and enforcement of access rights and restrictions

Sets of Records

- Persistent Sets
  - Files, subseries, series
- Lifecycle Management Sets
  - Record Groups, Digital Databases, Archival Microfilm Copies
- Transaction Sets
  - The records involved in a single transaction, such as a transfer, an accession, a response to a reference request, or mandatory review

Controls on Sets of Records

- Identify Sets, their members, and the order of members in the Sets
- Enable respect for original order
- Enable reconstruction of sets on successive generations of IT
- Set limits and criteria for transactions

Transactions Involving Records

- Transfer of physical custody
- Accessioning
- Conservation actions
- Access
- Review/Redaction
- Publication
- Destruction

Controls on Transactions Involving Records

- Ensure they are carried out in accordance with policies and sound archival and records management principles

- Enable automation of execution

- Determine success, failure and identify exceptions