The InterPARES Model for Preserving Authentic Electronic Records

William E. Underwood

Archiving International Standards

CCSDS Panel 2 Workshop

NASA Ames

May 14, 2001

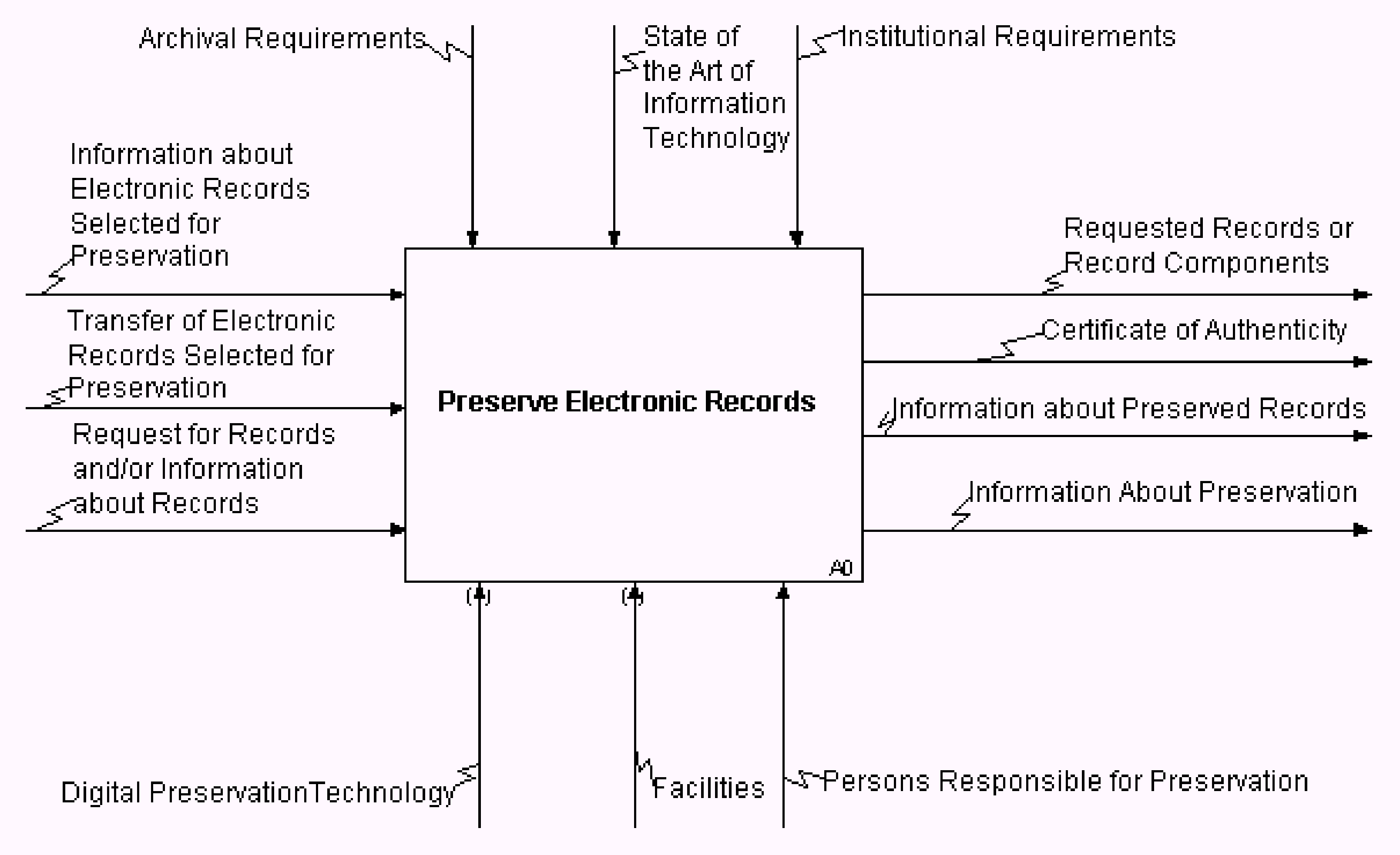
OVERVIEW

- The InterPARES Project
- The Preservation Task Force Model of “Preserve Electronic Records”
- A Set Theoretic Model for Preserving Authentic Digital Records

The InterPARES Project

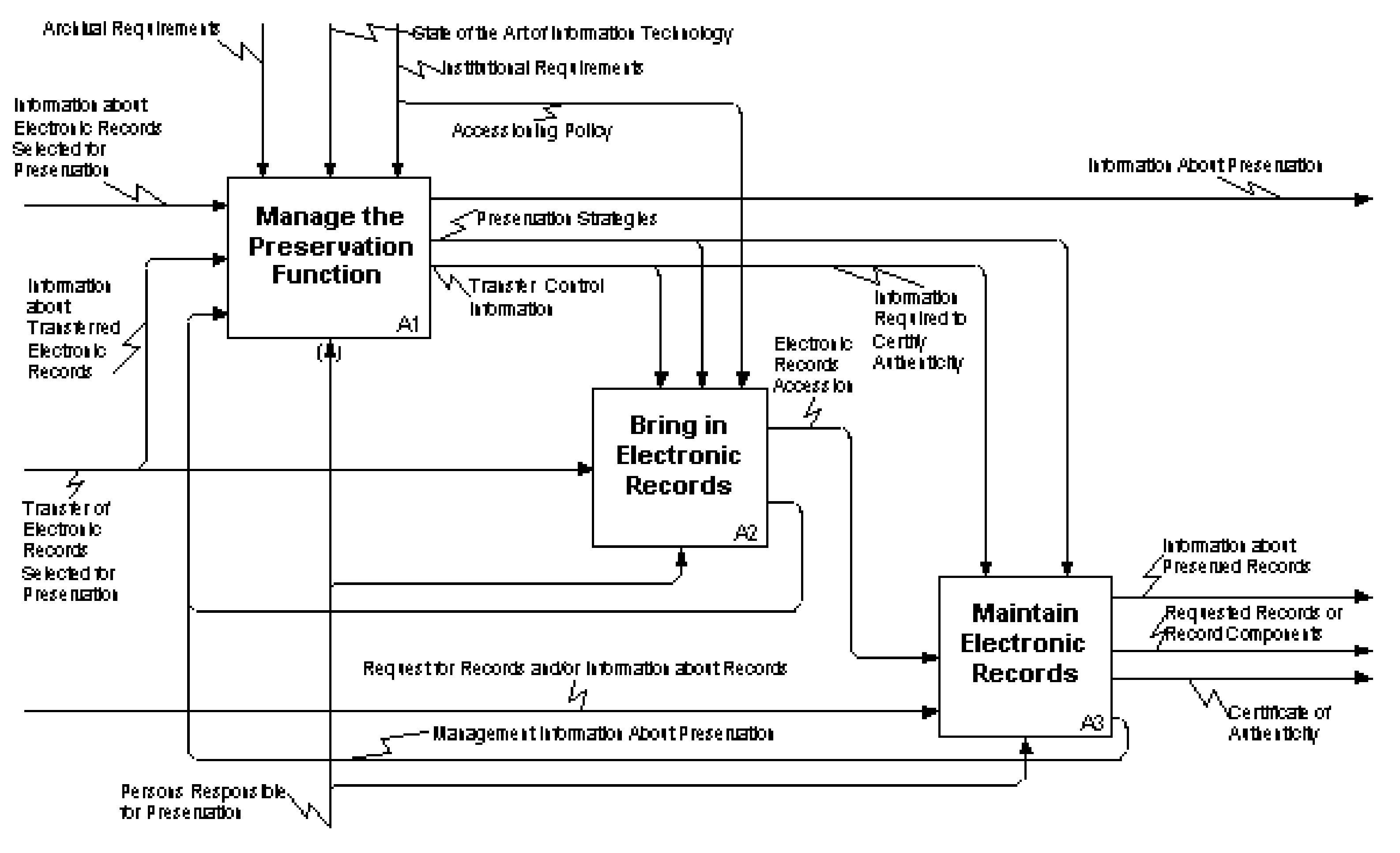
- Objectives:
  - develop guidelines for identifying requirements for preserving authentic records of long-term value, and
  - identifying technologies and procedures that support these requirements.
- International Team
  - National Archivists, Archival Scientists and a few Computer Scientists from 15 Nations.
- Task Forces
  - Authenticity, Appraisal, Preservation and Domain IV

USED AT: Vancouver Workshop #7	AUTHOR: Preservation Task Force PROJECT: InterPARES Project	DATE: 2/16/2000 REV: 2/16/2001	■	WORKING	READER	DATE	CONTEXT:
				DRAFT			TOP
				RECOMMENDED			
	NOTES: 1 2 3 4 5 6 7 8 9 10			PUBLICATION			


Archival Requirements
State of the Art of Information Technology
Institutional Requirements
Information about Electronic Records Selected for Preservation
Transfer of Electronic Records Selected for Preservation
Request for Records and/or Information about Records
Preserve Electronic Records
A0
Requested Records or Record Components
Certificate of Authenticity
Information about Preserved Records
Information About Preservation
Digital PreservationTechnology
Facilities
Persons Responsible for Preservation

NODE: A-0	TITLE: Preserve Electronic Records	NUMBER: 4.0

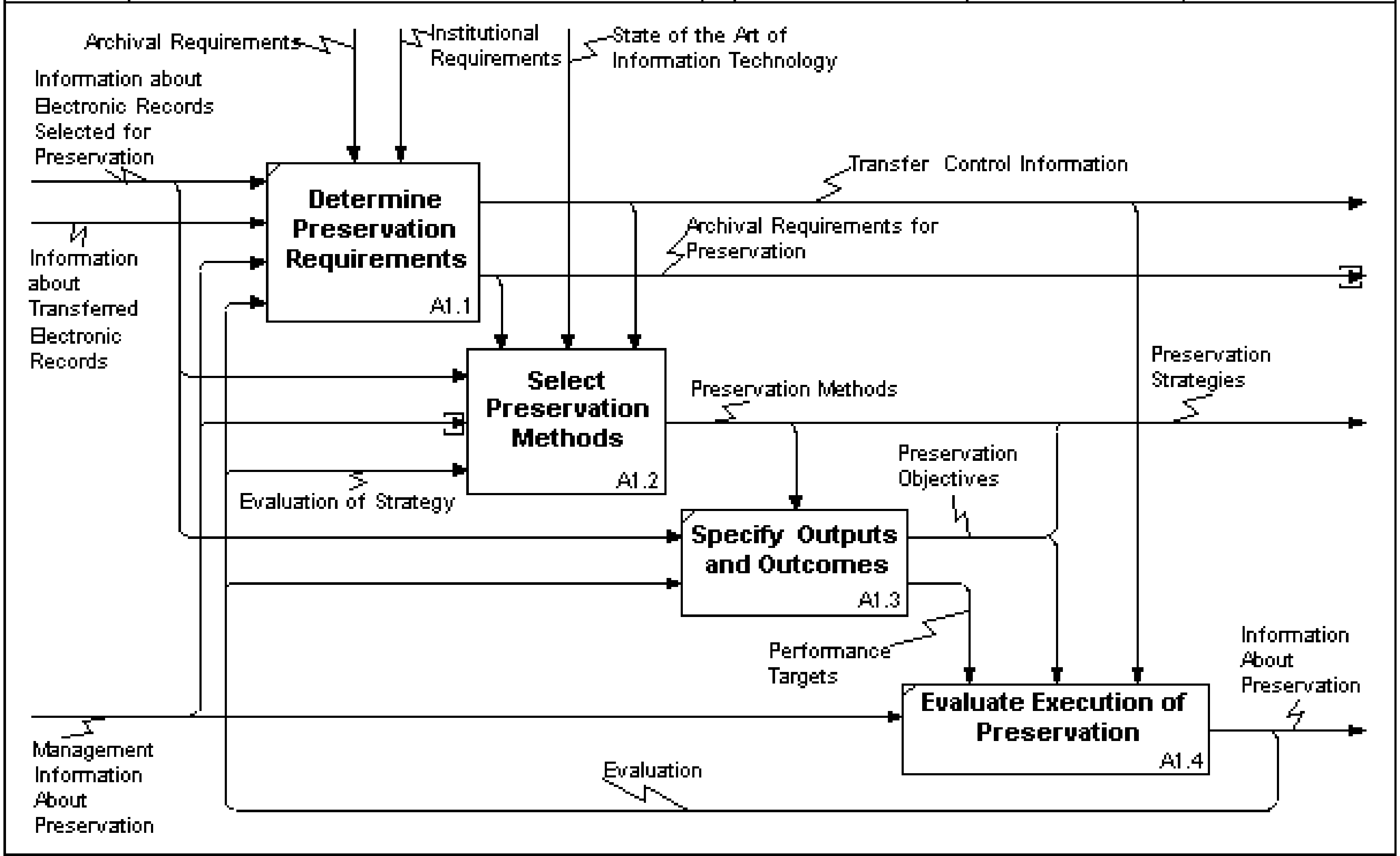
USED AT: Vancouver Workshop #7	AUTHOR: Preservation Task Force DATE: 2/17/2000 PROJECT: InterPARES Project REV: 2/15/2001 NOTES: 1 2 3 4 5 6 7 8 9 10	■ WORKING □ DRAFT □ RECOMMENDED □ PUBLICATION	READER DATE	CONTEXT: A-0



Archival Requirements
State of the Art of Information Technology
Institutional Requirements
Information about Electronic Records Selected for Preservation
Accessioning Policy
Information About Preservation
Manage the Preservation Function
A1
Preservation Strategies
Information about Transferred Electronic Records
Transfer Control Information
(A)
Information Required to Certify Authenticity
Electronic Records Accession
Bring in Electronic Records
A2
Transfer of Electronic Records Selected for Preservation
Information about Preserved Records
Maintain Electronic Records
A3
Requested Records or Record Components
Request for Records and/or Information about Records
Certificate of Authenticity
Management Information About Preservation
Persons Responsible for Preservation

NODE: A0	TITLE: Preserve Electronic Records	NUMBER: V 4.0

USED AT: Vancouver Workshop #7	AUTHOR: Preservation Task Force PROJECT: InterPARES Project NOTES: 1 2 3 4 5 6 7 8 9 10	DATE: 2/18/2000 REV: 2/19/2001	■ WORKING □ DRAFT □ RECOMMENDED □ PUBLICATION	READER DATE	CONTEXT: A0



Archival Requirements

Institutional Requirements

State of the Art of Information Technology

Information about Electronic Records Selected for Preservation

Information about Transferred Electronic Records

Determine Preservation Requirements A1.1

Transfer Control Information

Archival Requirements for Preservation

Select Preservation Methods A1.2

Preservation Methods

Preservation Strategies

Evaluation of Strategy

Preservation Objectives

Specify Outputs and Outcomes A1.3

Performance Targets

Evaluate Execution of Preservation A1.4

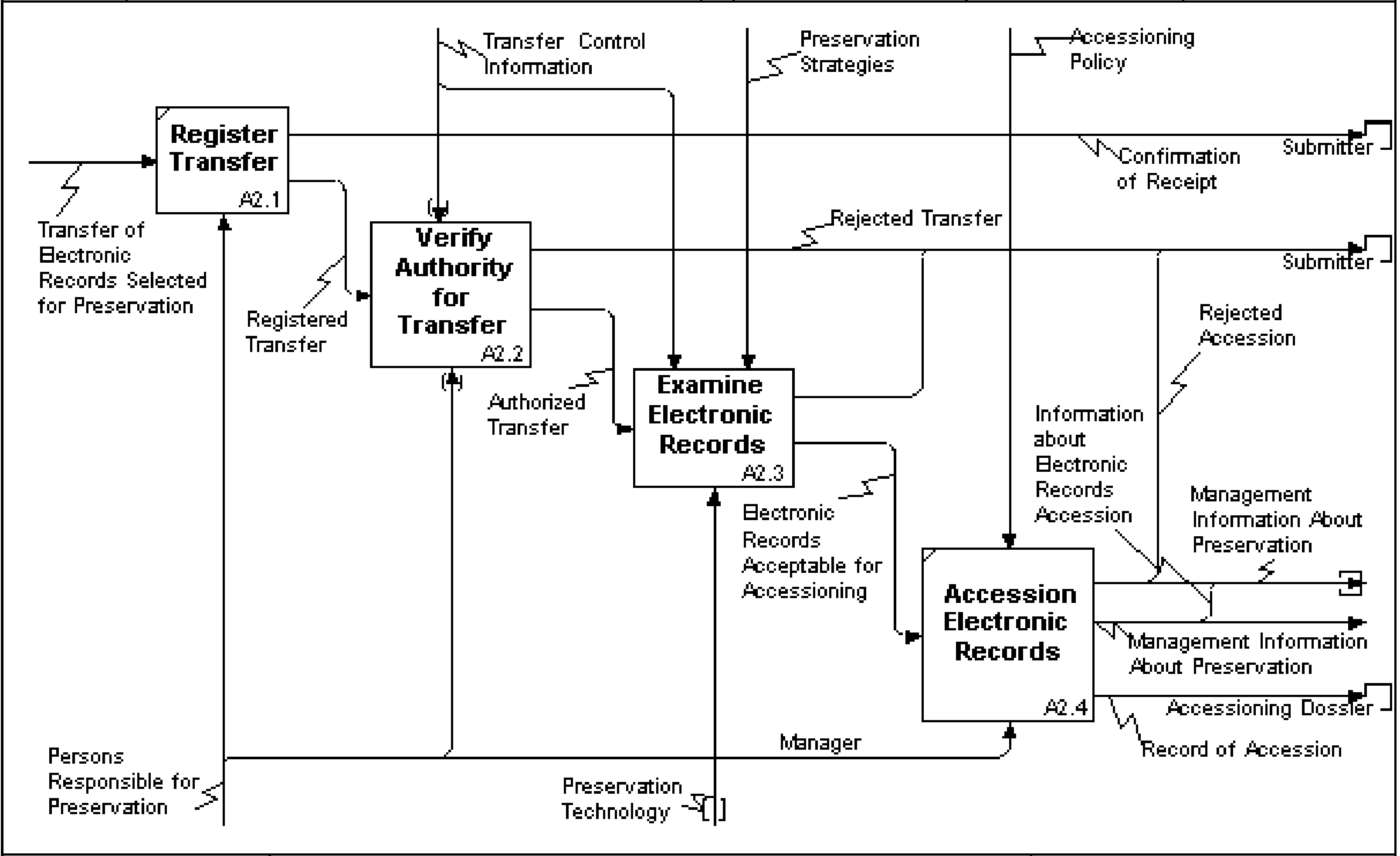
Information About Preservation

Management Information About Preservation

Evaluation

NODE: A1	TITLE: Manage the Preservation Function	NUMBER: V 4.0

USED AT:	AUTHOR: Preservation Task Force PROJECT: InterPARES Project NOTES: 1 2 3 4 5 6 7 8 9 10	DATE: 2/18/2000 REV: 6/25/2000	■ WORKING □ DRAFT □ RECOMMENDED □ PUBLICATION	READER DATE	CONTEXT: A0



Register Transfer
A2.1

Verify Authority for Transfer
A2.2

Examine Electronic Records
A2.3

Accession Electronic Records
A2.4

Transfer of Electronic Records Selected for Preservation

Transfer Control Information

Preservation Strategies

Accessioning Policy

Submitter

Confirmation of Receipt

Rejected Transfer

Submitter

Registered Transfer

Authorized Transfer

Rejected Accession

Information about Electronic Records Accession

Electronic Records Acceptable for Accessioning

Management Information About Preservation

Management Information About Preservation

Accessioning Dossier

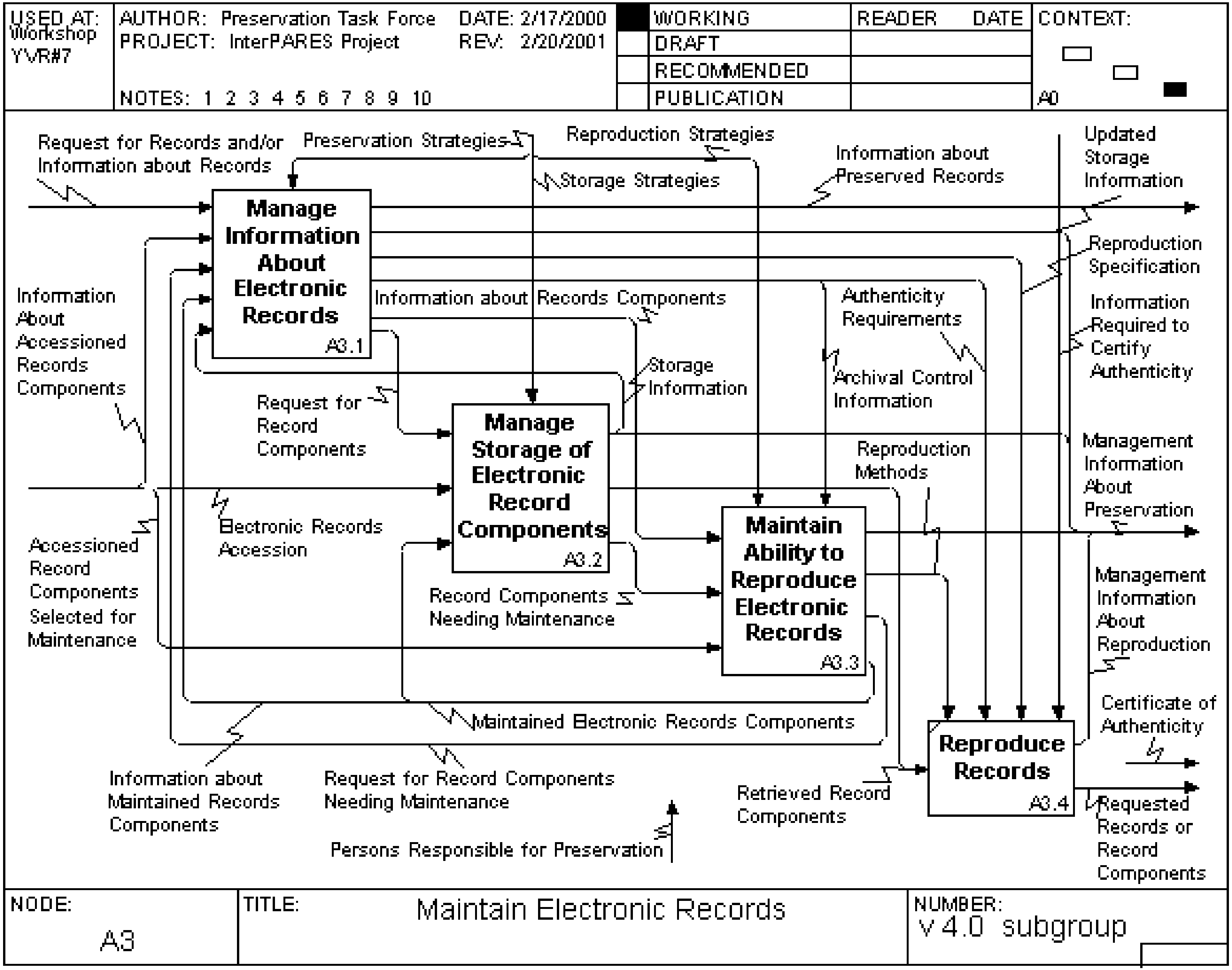
Record of Accession

Manager

Persons Responsible for Preservation

Preservation Technology

NODE:	TITLE:	NUMBER:
A2	Bring in Electronic Records	v. 4.0 subgroup



USED AT: Workshop YVR#7	AUTHOR: Preservation Task Force PROJECT: InterPARES Project NOTES: 1 2 3 4 5 6 7 8 9 10	DATE: 2/17/2000 REV: 2/20/2001	■ WORKING DRAFT RECOMMENDED PUBLICATION	READER DATE	CONTEXT: A0

Manage Information About Electronic Records A3.1

Manage Storage of Electronic Record Components A3.2

Maintain Ability to Reproduce Electronic Records A3.3

Reproduce Records A3.4

Request for Records and/or Information about Records

Preservation Strategies

Reproduction Strategies

Storage Strategies

Information about Preserved Records

Updated Storage Information

Reproduction Specification

Information Required to Certify Authenticity

Information About Accessioned Records Components

Information about Records Components

Authenticity Requirements

Storage Information

Archival Control Information

Request for Record Components

Reproduction Methods

Management Information About Preservation

Electronic Records Accession

Accessioned Record Components Selected for Maintenance

Record Components Needing Maintenance

Management Information About Reproduction

Certificate of Authenticity

Maintained Electronic Records Components

Information about Maintained Records Components

Request for Record Components Needing Maintenance

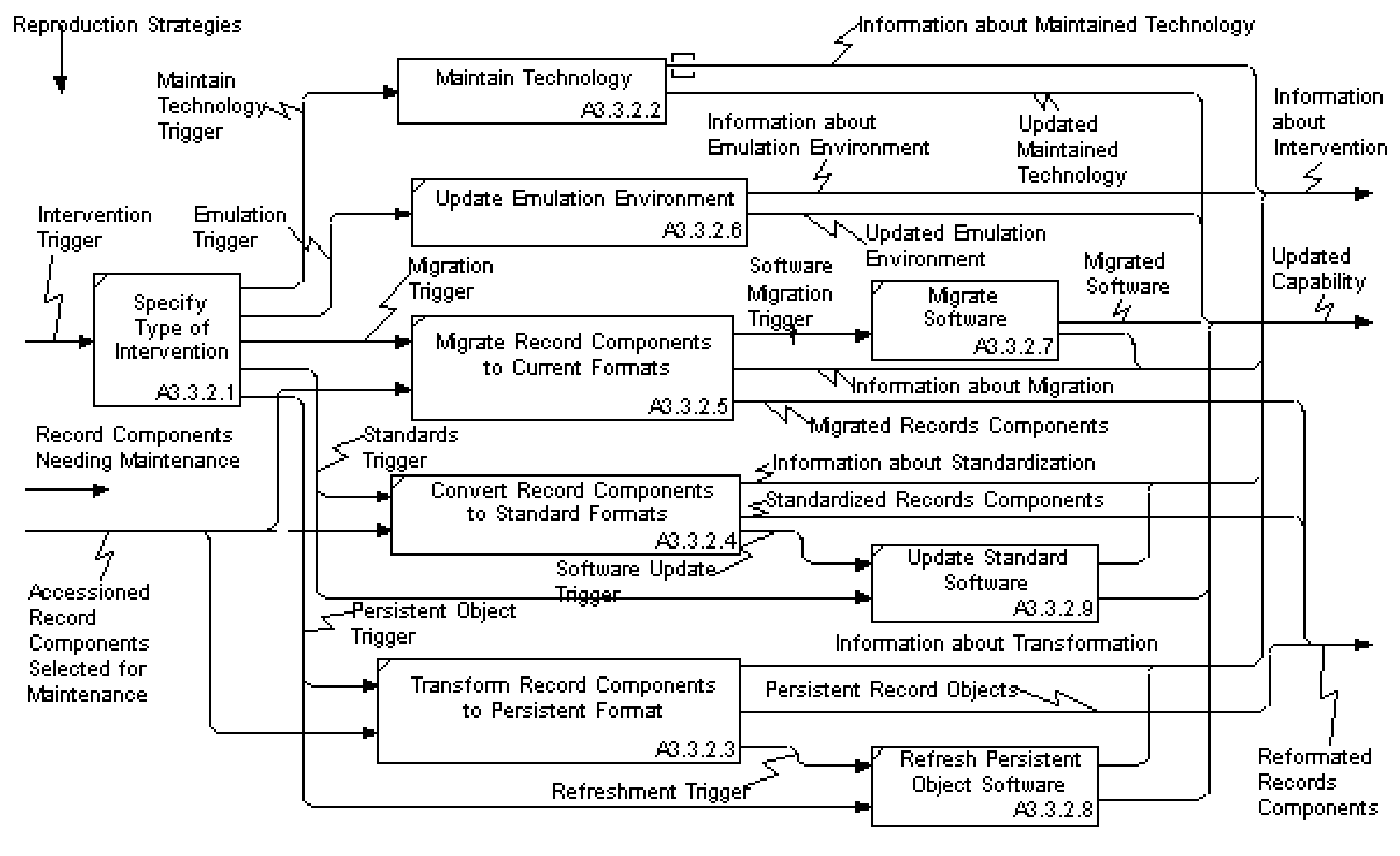
Retrieved Record Components

Requested Records or Record Components

Persons Responsible for Preservation

NODE: A3	TITLE: Maintain Electronic Records	NUMBER: V 4.0 subgroup

USED AT: Vancouver Workshop #7	AUTHOR: Preservation Task Force PROJECT: InterPARES Project NOTES: 1 2 3 4 5 6 7 8 9 10	DATE: 7/25/2000 REV: 2/19/2001	■ WORKING □ DRAFT □ RECOMMENDED □ PUBLICATION	READER DATE	CONTEXT: A3.3



Reproduction Strategies

Maintain Technology Trigger

Maintain Technology A3.3.2.2

Information about Maintained Technology

Information about Emulation Environment

Updated Maintained Technology

Information about Intervention

Intervention Trigger

Emulation Trigger

Update Emulation Environment A3.3.2.6

Updated Emulation Environment

Specify Type of Intervention A3.3.2.1

Migration Trigger

Software Migration Trigger

Migrate Software A3.3.2.7

Migrated Software

Updated Capability

Migrate Record Components to Current Formats A3.3.2.5

Information about Migration

Migrated Records Components

Record Components Needing Maintenance

Standards Trigger

Information about Standardization

Convert Record Components to Standard Formats A3.3.2.4

Standardized Records Components

Software Update Trigger

Update Standard Software A3.3.2.9

Accessioned Record Components Selected for Maintenance

Persistent Object Trigger

Information about Transformation

Transform Record Components to Persistent Format A3.3.2.3

Persistent Record Objects

Refreshment Trigger

Refresh Persistent Object Software A3.3.2.8

Reformatted Records Components

NODE: A3.3.2	TITLE: Execute Preservation Intervention	NUMBER: V 4.0

Set Theoretic Foundations of Digital Record Authenticity

- A *record* is a document made or received and set aside in the course of a practical business activity. [InterPARES Glossary]

- Record = {*x*: *x* ∈ Document and *O* ∈ Organization and *P* ∈ Persons and *P* ∈ *O* and *x* is made by *P* or *x* is received by *P* and *x* is saved in a filing system as evidence of a business activity of *O*}

Definition of Record Integrity

"The integrity of a record is its wholeness and soundness." [InterPARES Project, Draft Requirements for Authenticity]

Definition: *Digital record integrity* is the property of a record whereby the content and form of the record have not been altered in an unauthorized manner since the time the record was created, transmitted, or stored by an authorized source.

Definition of Authentic Record

"In order to verify the authenticity of a record, one must be able to verify its identity and its integrity. The identity of a record is provided by its provenance, author, addressee, writer, date, matter or action, and archival bond."

InterPARES Project, Draft Requirements for Authenticity, Nov 2000

- **Definition:** A digital record x is *authentic* iff P is a person who is a member of organization O and P created x at time d and x has not been altered in an unauthorized manner since time d.

Definition of authentic digital record series

- “A record series is file unit or documents arranged in accordance with a filing system or maintained as a unit because they result from the same accumulation or filing process, the same function, or the same activity.”
- A digital record series *S* is *authentic* iff all digital records *x* $\in$ *S* are authentic and the arrangement (file structure) of the records in the record series has not been altered since the time of creation of the records.

Java ARchive (JAR) File Technology

- JAR is a platform-independent file format that aggregates many files into one.
- JAR was developed so that Java applets and their components could be bundled into a single file and quickly downloaded to a browser in an http transaction.
- It provides the capability to verify the origin of components so that only those programs authored by those the user trusts will be executed.
- JAR is an open industry standard.

Preserving Files in a JAR

1. Create a JAR file that contains the files of a record series and a manifest file that contains the path/filenames of the files.
2. Create a message digest for each file and in the manifest file associate it with the path/filename of the file.
3. In the manifest file, associate the name of the record creator and archival date of each file with its path/filename.

Preserving files in a JAR (continued)

4. Create a message digest for the entire manifest file (the message digests of each of the files in the JAR and any metadata stored with the message digests) and store it in the signature file.

5. Sign the JAR file using an archival private key and the message digest for the manifest file. Insert the archival public key certificate file in the META-INF directory.

View of files in bundle.jar

META-INF/manifest.mf

META-INF/signature.sf

META-INF/signature.rsa

wp/corr/file1.wp5

wp/corr/file2.wp5

lotus/schedule.wks

lotus/budget.wks

photo/image1.jpg

photo/image.gif

Manifest File

```
XS = Manifest-Version 1.0
<creator
   organization = “Executive Office of the President”
   organizational-unit = “OPD”
<series
   title= "Richard Breeden's Files"
<folder
   title = "Alpha Correspondence 2-92"
<file
   id = "wp/corr/file1.wp5”
   sha1-digest = "TD1GZt8G11dXY2p40lSZPc5Rj64="/>
   format = "wp5.1"
   document-type = ”memo"
   name of author = "Breeden, Richard”
   name of creator = “Breeden, Richard”
   name of addressee = "Kristol, W; Kolb,
   archival-date = "01/12/92"…
```

Verifying the Integrity of Preserved Files

1. Extract the files from a JAR.

2. To ensure that the files in the JAR file haven’t changed since the JAR was signed, the message digests of each of the record files in the JAR are re-computed and compared with the message digests in the manifest.

3. The message digest for the message digests in the manifest are re-computed and compared against the message digest in the signature file.

4. Use the public key in the certificate in the Signature File to verify that the digital signature applied to the Manifest is that of an archival authority of the record creator’s organization.

Proving the Correctness of the Preservation Procedures

1. Express the assumptions and goals of the communication protocols and preservation procedures in a logical language

2. Make assertions in the logical language as to what is true after the execution of each procedural step.

3. Apply the set of axioms, definitions and deduction rules to the assumptions and results of procedural steps to derive the authentication goals.

Verification of Authenticity

Theorem: If a digital record series is believed to be authentic and is stored in a JAR that is digitally signed by an authorized member of the record creating organization using their private archival key, then at any time in the future, if the hardware and software to open the JAR and view the files it contains still exist, and the media on which the JAR is written has been periodically refreshed, then it can be verified whether the record series extracted from the JAR is authentic.

Applicability

Using this method, authenticity can be verified for:

- Active records stored in JARS in a record-keeping system
- Semi-active Records stored in JARS in a record center
- A transfer of inactive electronic records to an archives.
- Electronic records stored in an archives
- Records distributed to persons requesting archival records

Further Information

- **www.interpares.org (July 2001)**
  - **Requirements for Ensuring the Authenticity of Electronic Records over Time**
  - **IDEF0 Model of the Process of Appraising Electronic Records**
  - **IDEF0 Model of the Process of Preserving Electronic Records**
- **perpos.gtri.gatech.edu (July 2001)**
  - **Set Theoretic Foundations of Digital Record Authenticity**