

Choices and Strategies for Digital Records: The InterPARES Preservation Model

William Underwood,
Georgia Tech Research Institute
Atlanta, Georgia, USA

Choices and Strategies for Preservation of the Collective Memory
Grand Hotel, Toblach/Dobbiaco 25-29 June 2002

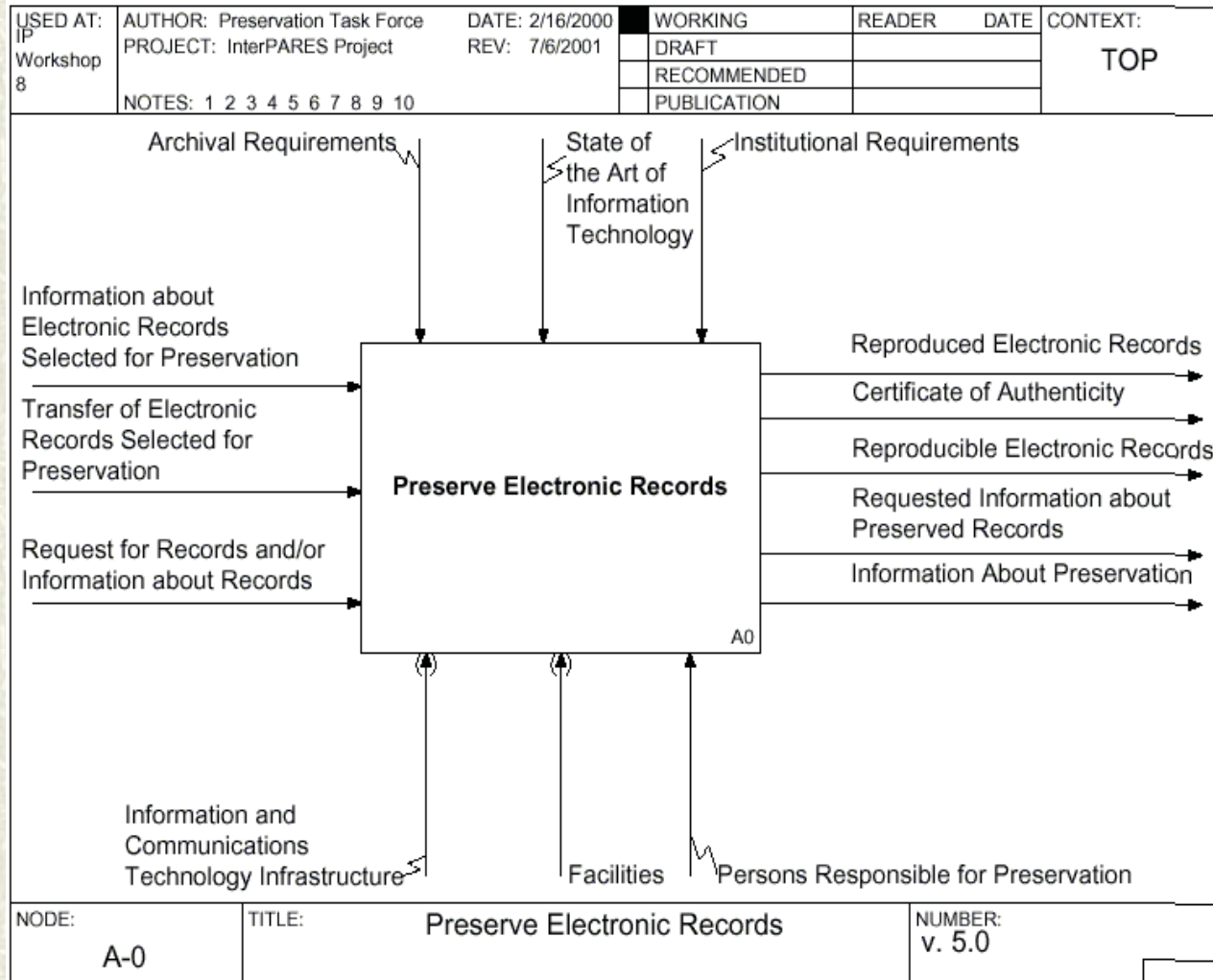
Overview

- # Analysis of the problem of preserving authentic electronic records and the resulting preservation model.
 - # Results of a walkthrough to test the model.
 - # Demonstrate that the Preservation Model satisfies the Baseline Requirements
 - # Conclusions & future research
-

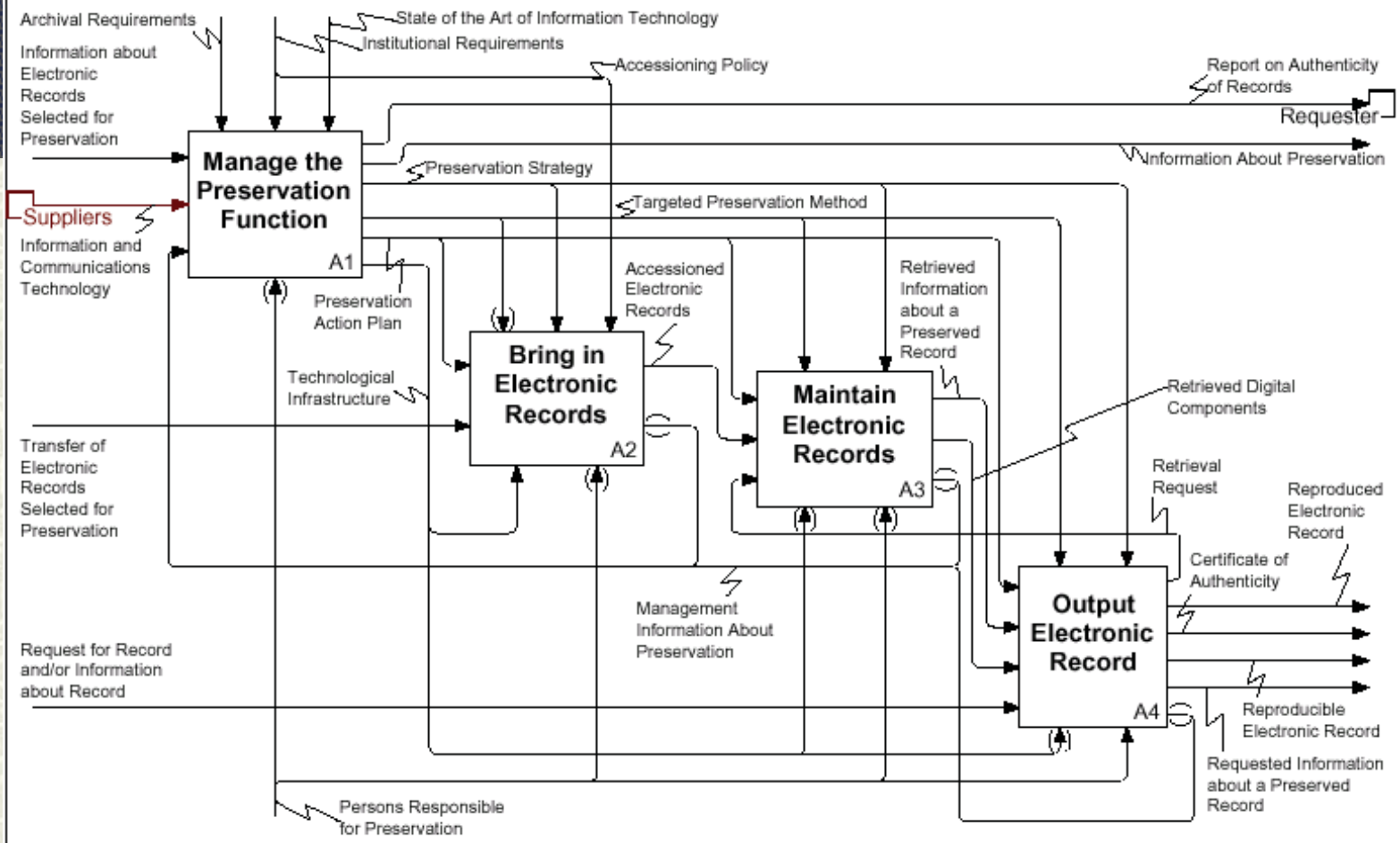
Research Objective & Approach

- # A generic solution to the problem of preserving authentic electronic records.
 - # Use the IDEF0 notation and methodology to represent the problem and the results of the analysis of the problem.
-

The Problem of Preserving Authentic Electronic Records

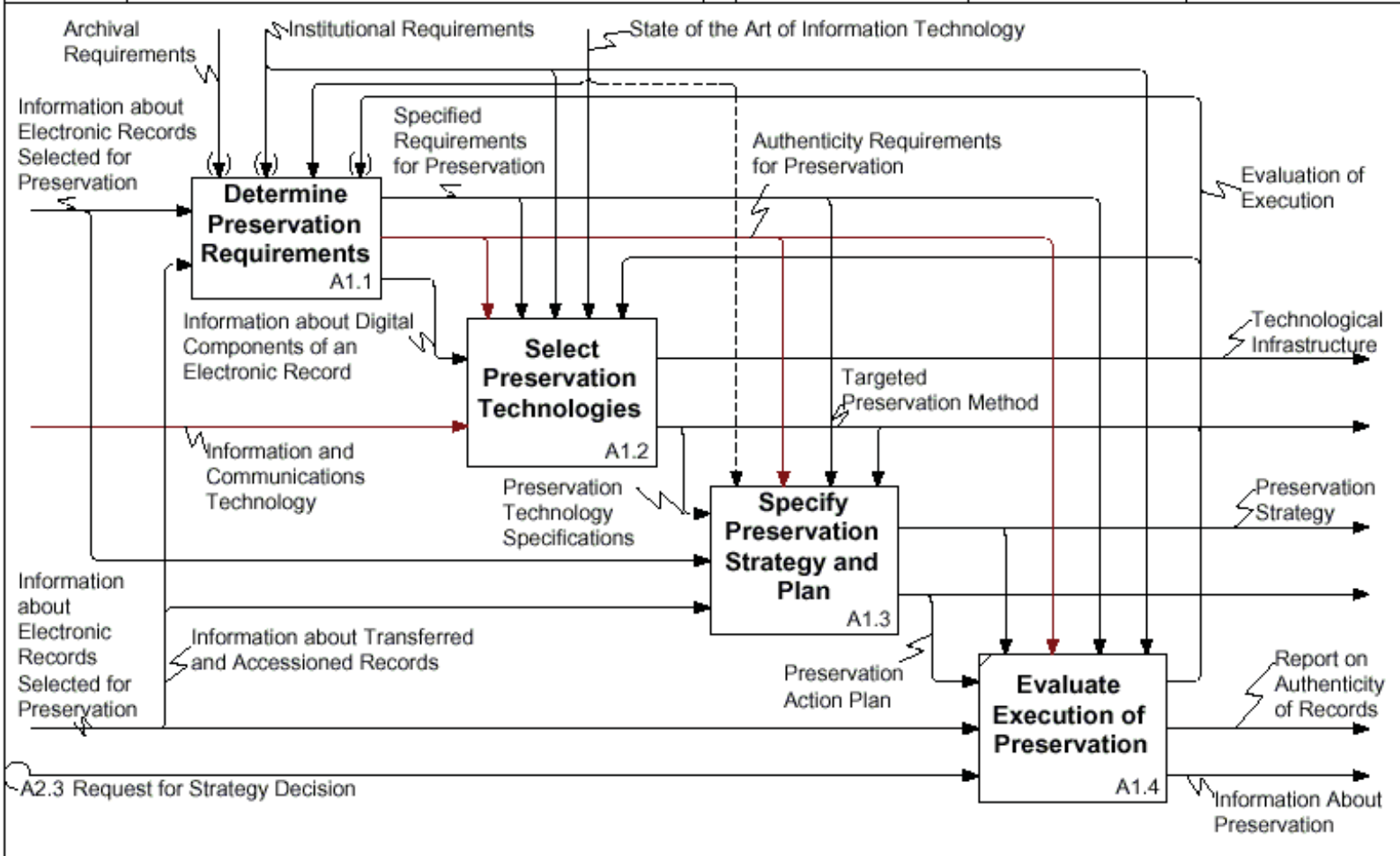


USED AT: PTF Workshop 8 2001	AUTHOR: Preservation Task Force	DATE: 2/17/2000	WORKING	READER	DATE	CONTEXT: A-0
	PROJECT: InterPARES Project	REV: 9/30/2001	DRAFT			
	NOTES: 1 2 3 4 5 6 7 8 9 10		RECOMMENDED			
			PUBLICATION			



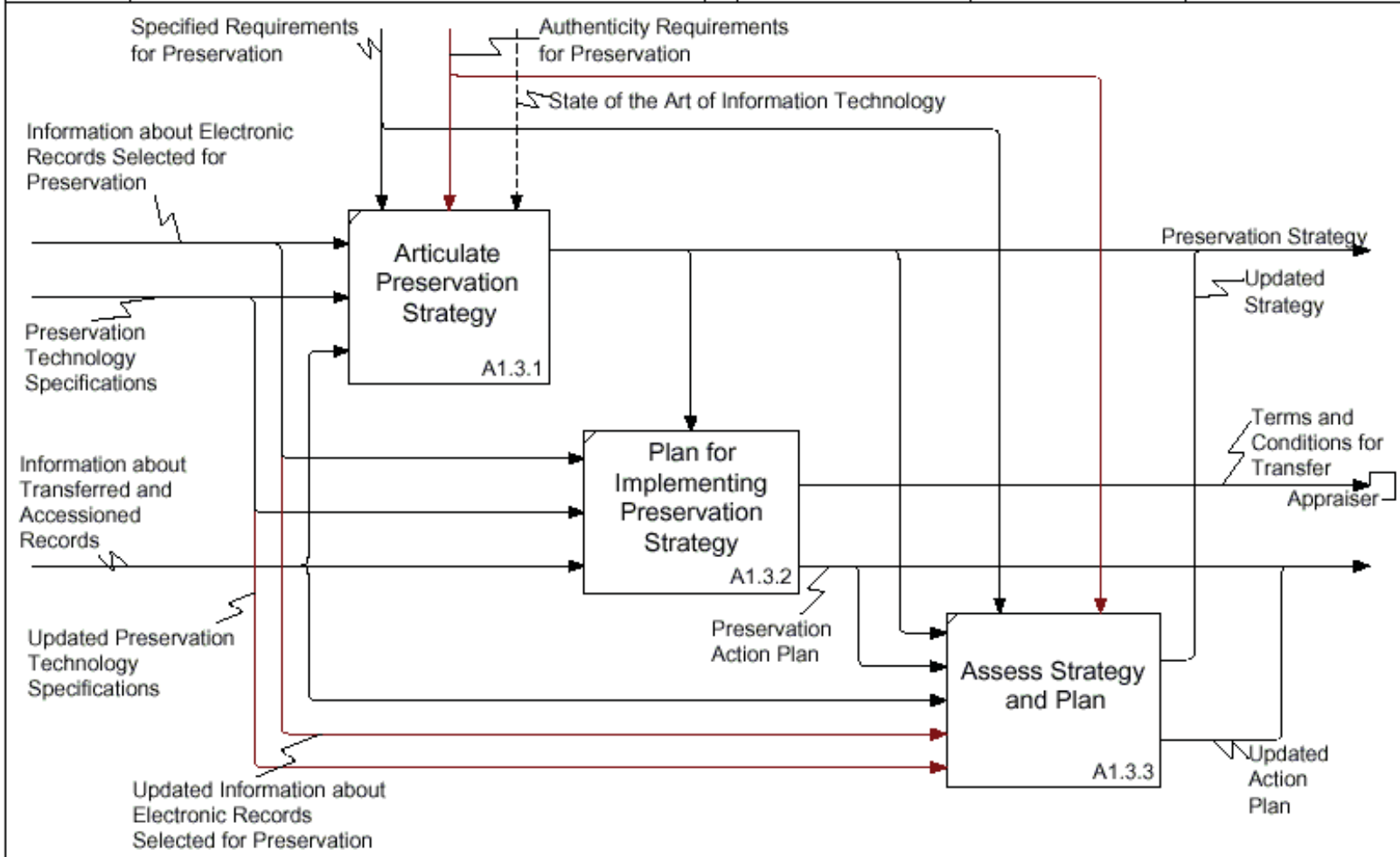
NODE: A0	TITLE: Preserve Electronic Records	NUMBER: v 5.1
--------------------	--	-------------------------

USED AT: PTF Workshop 8 2001	AUTHOR: Preservation Task Force	DATE: 2/18/2000	WORKING	READER	DATE	CONTEXT: A0
	PROJECT: InterPARES Project	REV: 9/30/2001	DRAFT			
	NOTES: 1 2 3 4 5 6 7 8 9 10		RECOMMENDED			
			PUBLICATION			



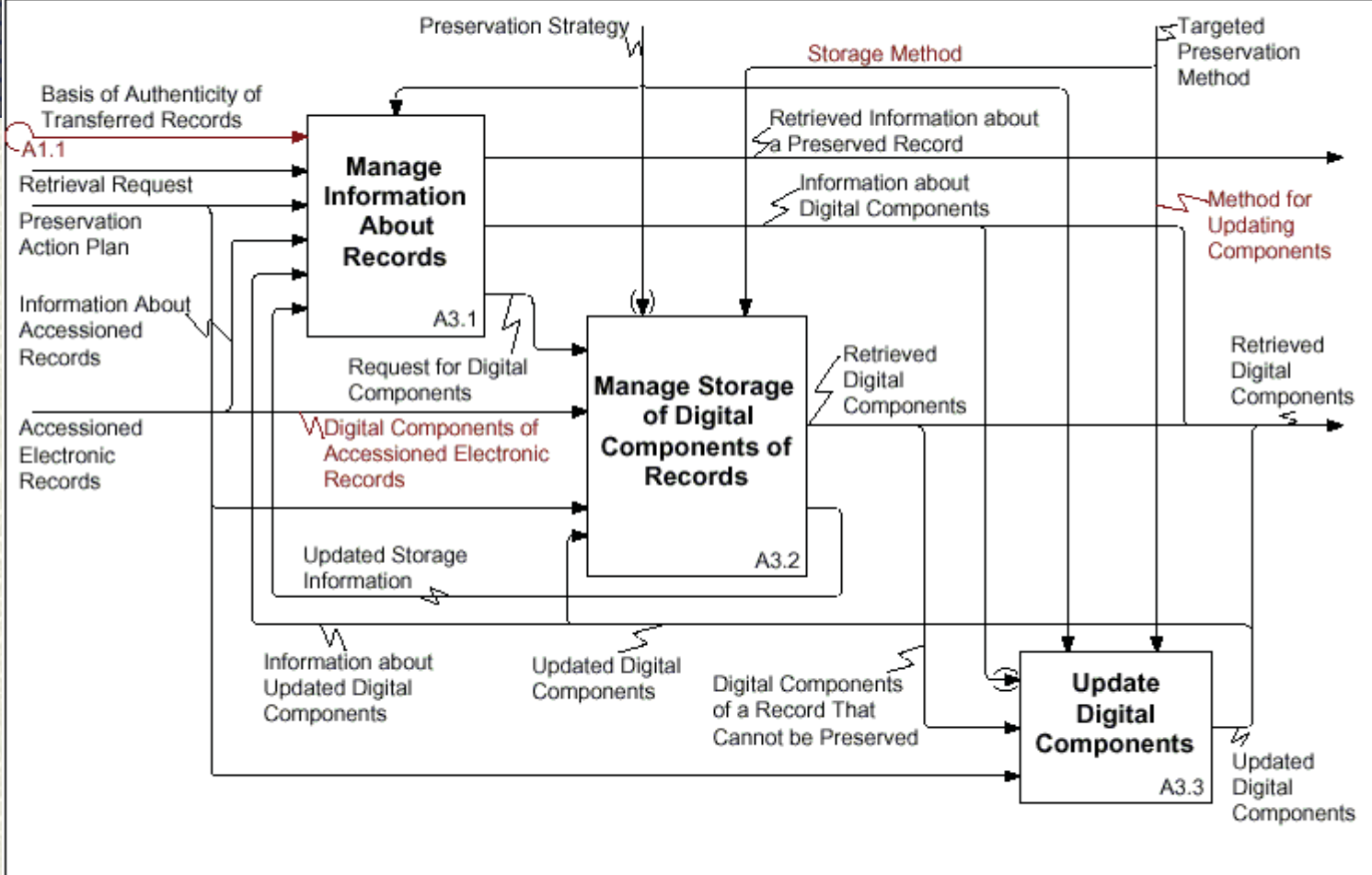
NODE: A1	TITLE: Manage the Preservation Function	NUMBER: v 5.1
--------------------	---	-------------------------

USED AT: PTF Workshop 8 2001	AUTHOR: Preservation Task Force	DATE: 5/11/2001	WORKING	READER	DATE	CONTEXT: A1
	PROJECT: InterPARES Project	REV: 10/6/2001	DRAFT			
	NOTES: 1 2 3 4 5 6 7 8 9 10		RECOMMENDED			
			PUBLICATION			



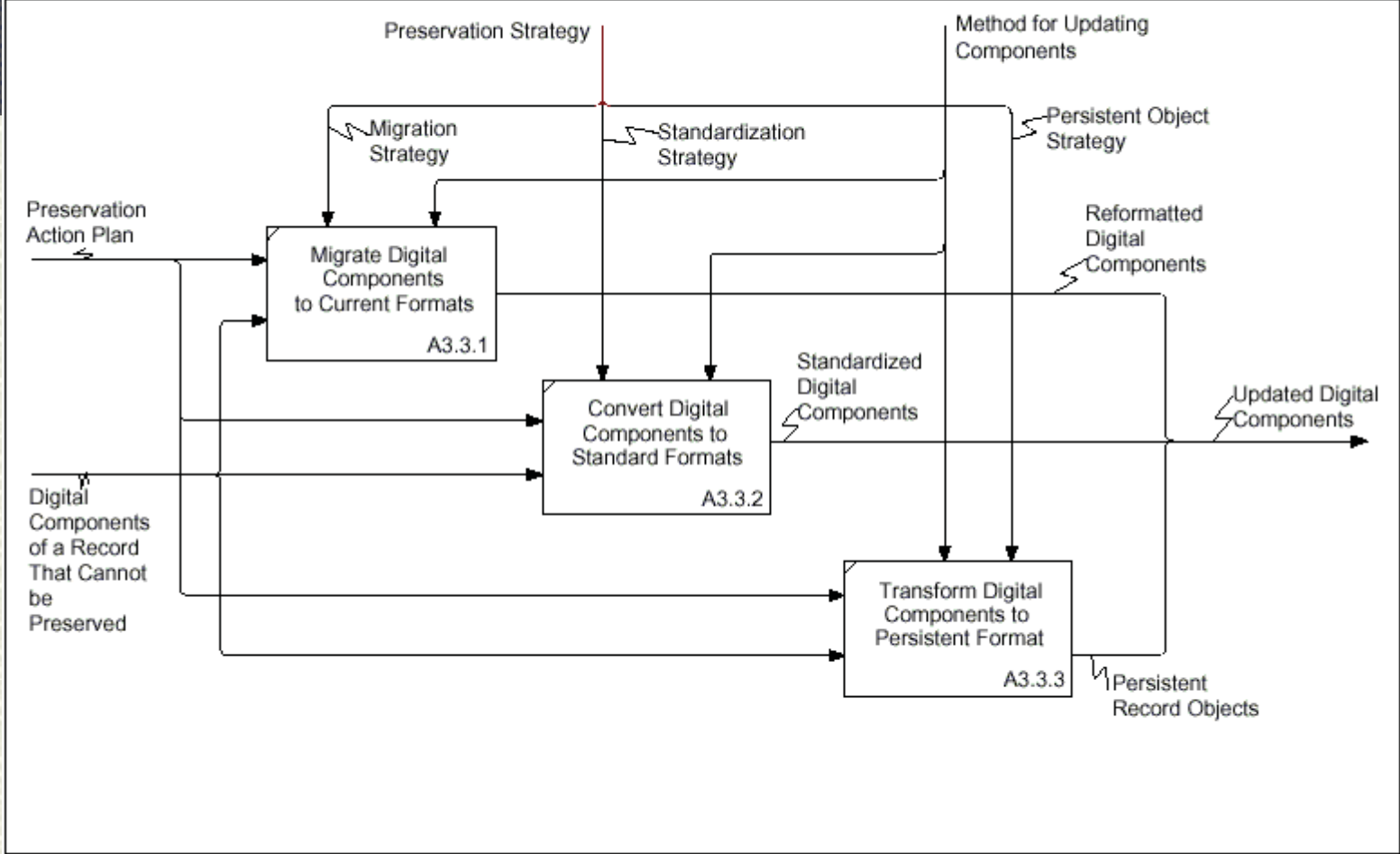
NODE: A1.3	TITLE: Specify Preservation Strategy and Plan	NUMBER: v 5.1
----------------------	---	-------------------------

USED AT: PTF Workshop 8 2001	AUTHOR: Preservation Task Force	DATE: 2/17/2000	WORKING	READER	DATE	CONTEXT: <input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/>	
	PROJECT: InterPARES Project	REV: 9/30/2001	DRAFT				
	NOTES: 1 2 3 4 5 6 7 8 9 10			RECOMMENDED			A0
				PUBLICATION			



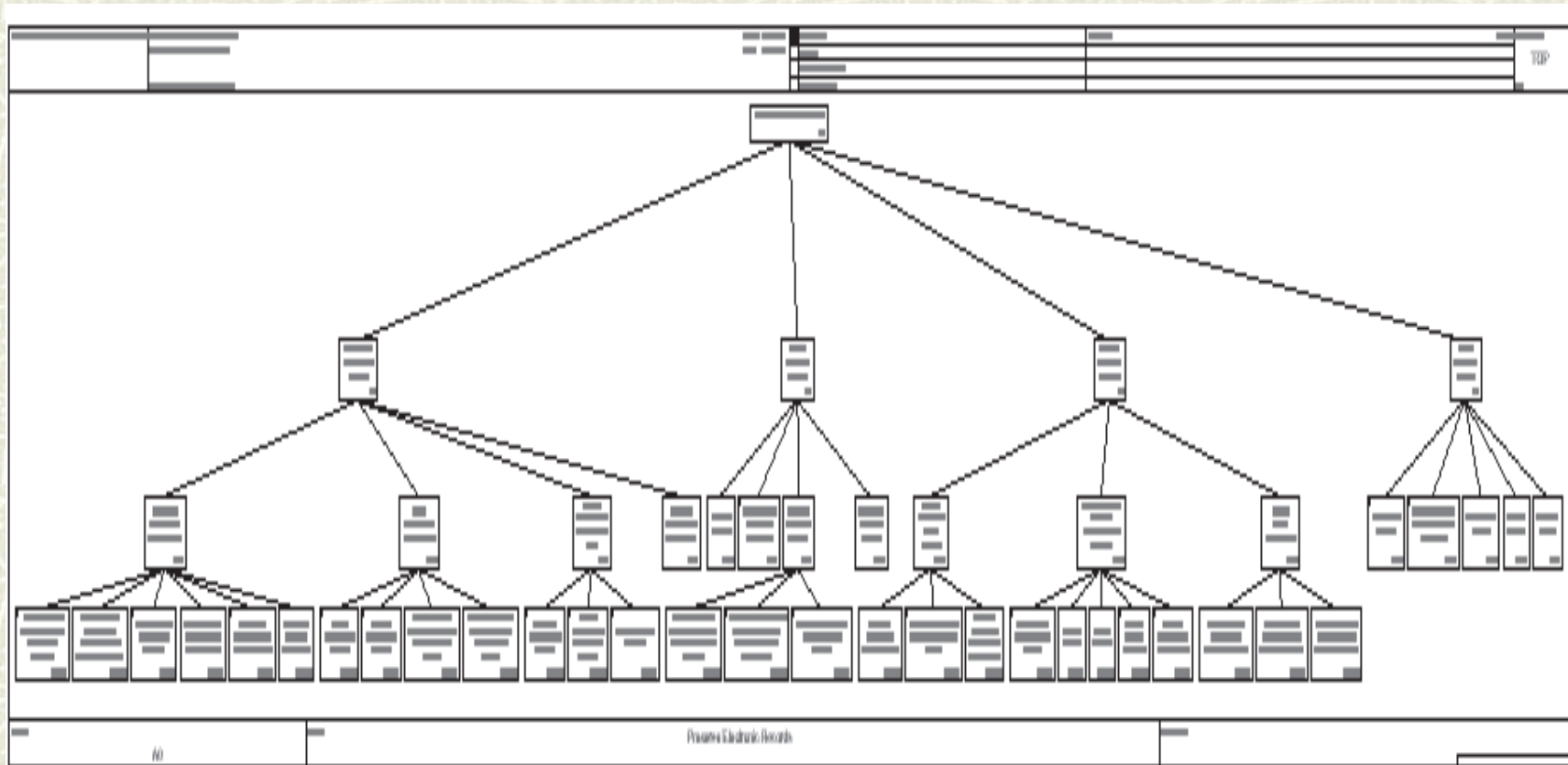
NODE: A3	TITLE: Maintain Electronic Records	NUMBER: v 5.1
--------------------	--	-------------------------

USED AT: PTF Workshop 8 2001	AUTHOR: Preservation Task Force	DATE: 7/11/2000	WORKING	READER	DATE	CONTEXT: <input type="checkbox"/> <input type="checkbox"/>
	PROJECT: InterPARES Project	REV: 10/6/2001	DRAFT			
	NOTES: 1 2 3 4 5 6 7 8 9 10		RECOMMENDED			A3
			PUBLICATION			



NODE: A3.3	TITLE: Update Digital Components	NUMBER: v 5.1
----------------------	--	-------------------------

The Problem Reduction Tree for the Preservation Problem



Walkthrough of the Model using Case Study Data

Walkthrough Objectives:

- To demonstrate that the preservation model works for actual cases of records selected for preservation.
- To test and refine, and eventually validate, the Preservation Model.
- To determine whether the Preservation Model satisfies the Baseline Requirements.

Case Study Data:

- New York State Worker's Compensation Board (NY WCB)
 - Electronic Case Folder System (ECFS)
-

Walkthrough Team

- # Presenter
 - # Reviewers
 - # Case Study Expert
 - # Secretary
-

The Preservation Framework

- # Terms and Conditions for Transfer
 - # Assessment of a Presumption of Authenticity
 - # Preservation Strategy for a Specific Body of Records.
 - # Preservation Action Plans
 - # Preservation Methods
-

Example of Terms and Conditions for Transfer

Record Creator's Name	NY State Workers' Compensation Board
Transfer Agent's Name	John Doe, Records Manager
Identification of Records Title: Description: Document Types: File Format: Volume: Data Structure:	NY State Electronic Case Folder System Series of case files for adjudicating benefits of disabled workers Claims for Benefits, Employer's reports of accidents and illness, Correspondence, Medical Reports, Insurance Carrier's Reports Multipage TIFF 300,000 cases, 1 million files, 9 million document images Relational Schema
Scheduled Transfer Date	TBD
Medium for Transfer	DLT Tape
Technical Conditions for Transfer	MD5 hash code of all files for integrity check

Information Supporting an Assessment of Authenticity

Benchmark Requirement	Information Supporting Presumption of Authenticity
<p>A.1.a Identity of the record</p> <p>A.1.a.i Name of author Name of addressee</p> <p>A.1.a.ii Name of action or matter</p> <p>A.1.a.iii Chronological date</p> <p>A.1.a.iv Expression of Archival Bond</p> <p>A.1.a.v Indication of attachments</p>	<p>The ECFS data model permits the association of author's name, addressee, name of action or matter, and chronological date with each document.</p> <p>When documents are imported by FileNet, Case File is ordered by document number. Document preparation and mail transmittal preparation rules address how attachments are kept in the case folder.</p>
<p>A.1.b Integrity of the record</p> <p>A.1.b.i Name of Handling Office</p> <p>A.1.b.ii Name of OPR</p> <p>A.1.b.iii Indication of types of annotations</p> <p>A.1.b.iv Indication of technical modifications</p>	<p>NY WCB</p> <p>Same as above</p> <p>FileNet Supports annotations, but they are not used.</p> <p>Paper documents are scanned into document images in TIFF 6 format and maintained on WORM disks.</p>

Assessment of Authenticity

A.2 Access Privileges	Access to ECFS is controlled via passwords, job titles, workgroups, geographic location and business need.
A.3 Protective Privileges: Loss and Corruption of Records	There are backup copies of the WORM disks and transaction logs.
A.4 Protective Privileges: Media and Technology	WORM Disks are guaranteed for over 100 years.
A.5 Establishment of Documentary Forms	Each form is described in a procedural manual that is managed in Lotus Notes.
A.6 Authentication of Records	Authentication of document images in a case file is occasionally required in the adjudication process. They are presumed authentic because they are scanned images of paper documents and they are used in the normal course of business.
A.7 Identification of Authoritative Record	The images are the authoritative record unless the paper file is still available.
A.8 Removal and Transfer of Relevant Documentation	There has not yet been a transition of active records to semi-active or inactive status, which involves the removal of records from the electronic system.

Examples of Digital Preservation Strategies

To overcome obsolescence of a computer platform,

- Migrate file viewers to new platform, or
 - Convert obsolete proprietary formats to current proprietary or standard formats, or
 - Convert proprietary formats to descriptions in a standard markup language, or
 - Emulate obsolete processors, storage and display devices on current processors, storage and display devices.
-

An Example of a Preservation Action Plan

1. Retrieve digital components for “Claims for Benefits” in Electronic Case Folder System (ECFS) that are ASCII Files.
 2. Convert the ASCII text files to TIFF multi-page format using TIFFmaker.
 3. Store the digital components converted to TIFF format back to Archival Storage.
 4. Store in the database “on *current date* the digital components for Claims for Benefits in the ECFS that were in ASCII format were converted to TIFF multi-page format.”
-

Examples of Preservation Methods

Method	Software
Check integrity of transferred records	Message Digest(MD5, SHA1)
Package Record Components	tar, untar, JAR
Storage Update Method	Tape Copy
Maintain Information about Records and Digital Components	DBMS (Oracle, Sybase)
Archival Storage	IBM High Performance Storage System
Reproduce Records	TIFF and PDF viewers, X86 emulator
Updating Components	TIFFmaker, Word2PDF, Word2XML

The Preservation Framework Satisfies the Baseline Requirements

Requirement	Satisfied by
1.Controls over Records Transfer, Maintenance, and Reproduction	Terms & Conditions of Transfer; Preservation Action Plans; Preservation & Reproduction Methods
1.a Unbroken Custody of the records is maintained	Information about Records Selected for Preservation; Transfer and Maintenance of Records.
1.b Security and control procedures are implemented and monitored	Access Control and Access Privileges of a DBMS
1.c The content of the record remains unchanged after reproduction	Selecting preservation methods that preserve content & verifying that records can be reproduced.

The Preservation Framework Satisfies the Baseline Requirements

Requirement	Satisfied by
2. Documentation of Reproduction Process and its Effects	The effect of selected reproduction methods is evaluated.

The Preservation Framework Satisfies the Baseline Requirements

Requirement	Satisfied by
3. The archival description of the fonds containing the electronic records includes information about changes the electronic records of the creator have undergone since they were first created.	Preservation Action Plans store records of refreshment of digital media and updates to digital components.

Conclusion: Significant progress has been made.

- # The InterPARES Preservation Model is a framework that archival institutions can use to develop strategies for preserving authentic electronic records.
 - # The model satisfies the Baseline Requirements.
 - # The walkthrough shows that the model works for real preservation problems.
 - # Metadata was identified during the walkthrough that can be used to create a preservation data model.
-

Conclusion: However, the model needs refinement.

- # Walkthrough identified a number of opportunities for improving the model.
 - # Validate the model with additional walkthroughs
 - # Need preservation data model
 - # Empirical results are needed in applying alternative preservation strategies and evaluating their relative cost effectiveness.
 - # Technical methods are needed for authenticating electronic records
-

Further Information

Preservation Task Force Final Report:

www.interpares.org

Empirical Studies of Preservation Strategies
for Personal Computer Records:

perpos.gtri.gatech.edu

Technical Methods for Authentication of
Digital Records:

is.gseis.ucla.edu/us-interpares/
