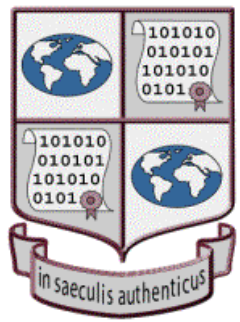
Preserving Electronic Records

The Work of the Preservation Task Force

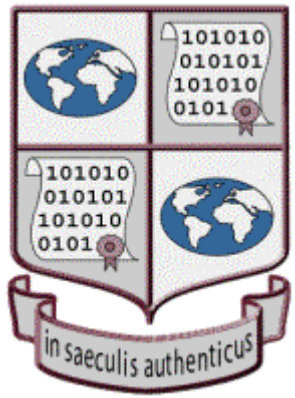
"There's Nothing Like the Real Thing"
Preserving Authentic Electronic Records:

The Findings of InterPARES I

June 19, 2002



InterPARES Project
Preservation Task Force



Kenneth Thibodeau, NARA, Chair
Richard Blake, Public Records Office, UK
Paola Caruci, National Archives of Italy
Michele Cloonan, University of California, Los Angeles
Babak Hamidzadeh, University of British Columbia
P.C. Hariharan, Johns Hopkins University
Hans Hofman, National Archives of the Netherlands
Torbjörn Hörnfeldt, National Archives of Sweden
Richard Lysakowski, Collaborative Electronic Notebooks Systems Assn.
Christine Petillat, National Archives of France
William Rhind, Pharmacia Corporation
William Underwood, Georgia Tech Research Institute
Bruce Walton, National Archives of Canada

Preserving Authentic Electronic Records

- What is required to preserve authentic records is an **archival question**.
- **Technology** provides – and limits – **options** for implementing the answer(s) to the archival question.
- Archival Requirements and Technological Solutions are melded together in a **Preservation Strategy** for a given body of records.
- Anyone responsible for preservation should develop a **Preservation Framework** both to ensure that its Preservation Strategies are coherent and to enable evolution of those strategies over time.

Preserving Electronic Records

- **It is impossible to preserve an electronic record.**
- **It is only possible to preserve the ability to reproduce an electronic record.**
  - Digital data inscribed on a physical medium do not have the form of a record.
  - It is necessary to transform inscribed bits into the form of the record.
  - The transformation is done by software.
  - An electronic record is reproduced by the correct processing of the stored sequence(s) of bits which encode not only the content, but also all the intrinsic and extrinsic elements of form of the record.

Reproducing an Electronic Record Correctly

- All the necessary sequences of stored bits must be retrieved without error.
- The right software must be used.
- The software must function properly.
- **If the reproduction has all the identifying characteristics of the record and its integrity has been maintained, it is an authentic copy of the record.**

Digital Component of an Electronic Record

- A digital object that is part of an electronic record, or of a reproduced electronic record, or that contains one or more electronic records, or reproduced electronic records, and that requires specific methods for preservation.

Digital Objects

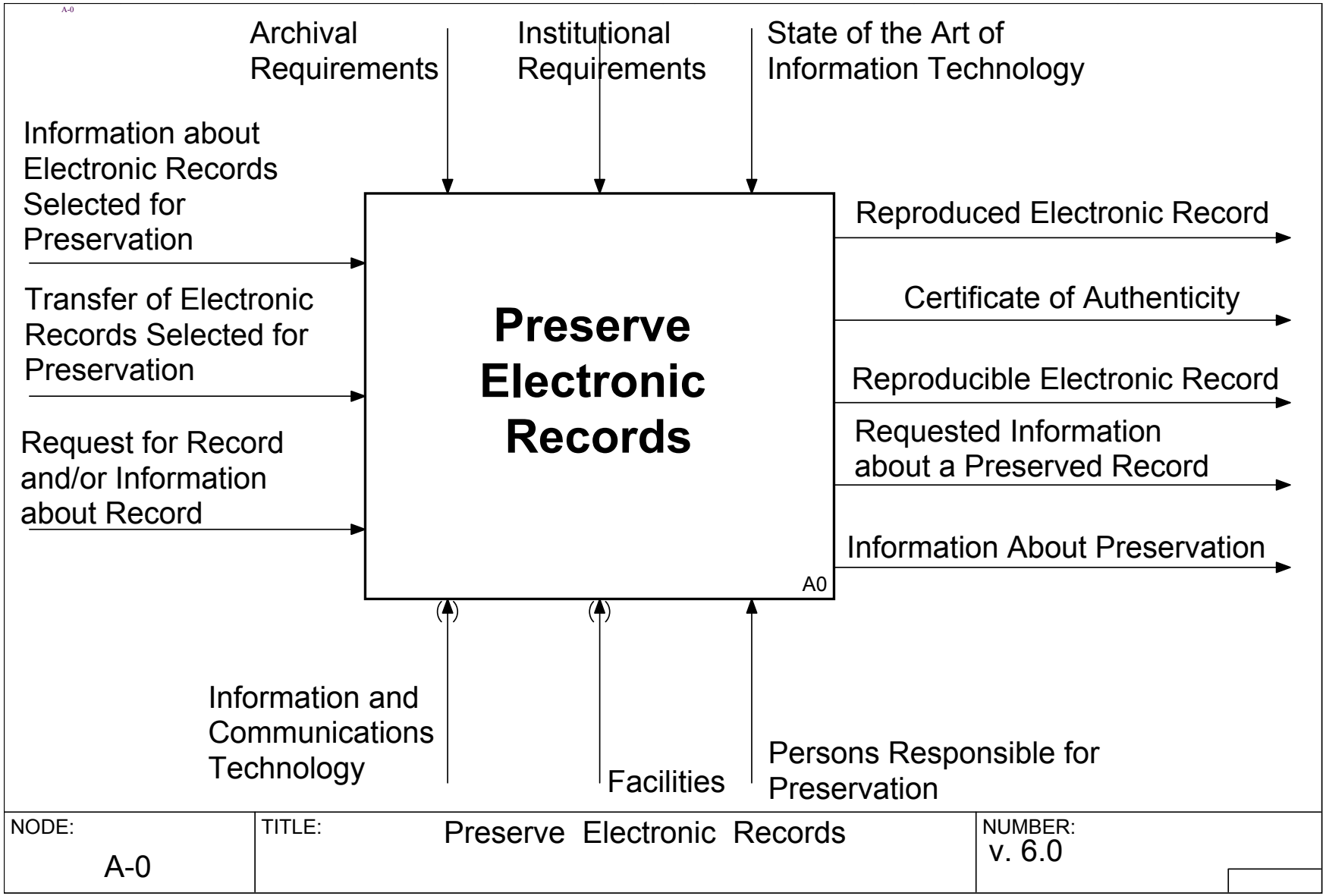
- Multiple Inheritance:
  - Physical Object
    - An inscription of signs on a physical medium
  - Logical Object
    - A digital object recognized (& processed) by software
  - Conceptual Object
    - The object as recognized and understood by a person.

Relationships of the Physical, Logical, & Conceptual Levels of an Object

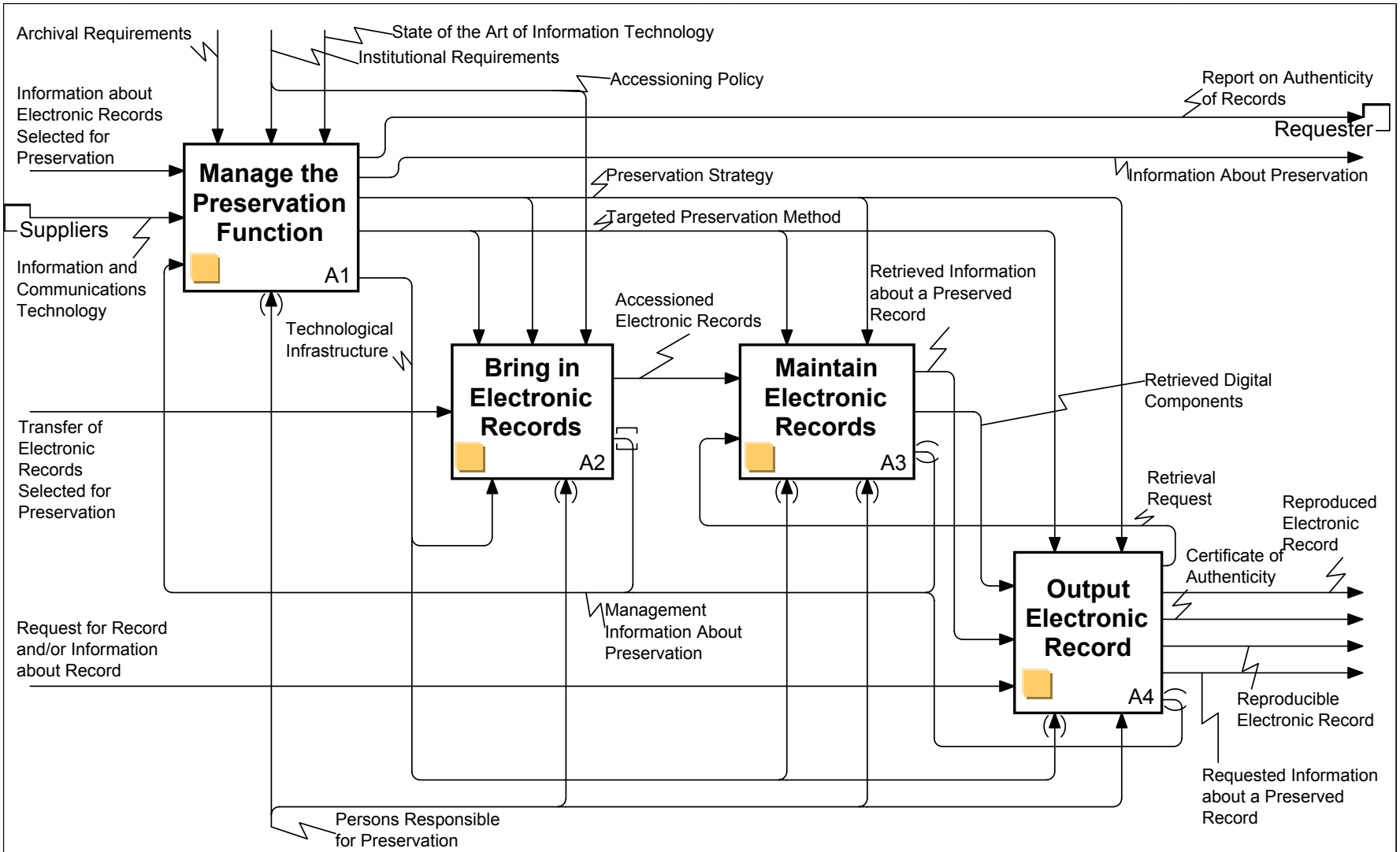
- One-to-one
  - A report created with a word processing application, saved as a word processing file, copied to diskette.
- One-to-many
  - A long report divided into a master and 3 subdocuments.
  - A digital photograph included in a textual report, but stored in a separate, linked jpeg file.
- Many-to-one
  - 200 word processing files stored in a TAR file.
- Many-to-many
  - Data elements from several different database tables combined, in different ways, to produce various reports.

The "Preserve Electronic Records" Model

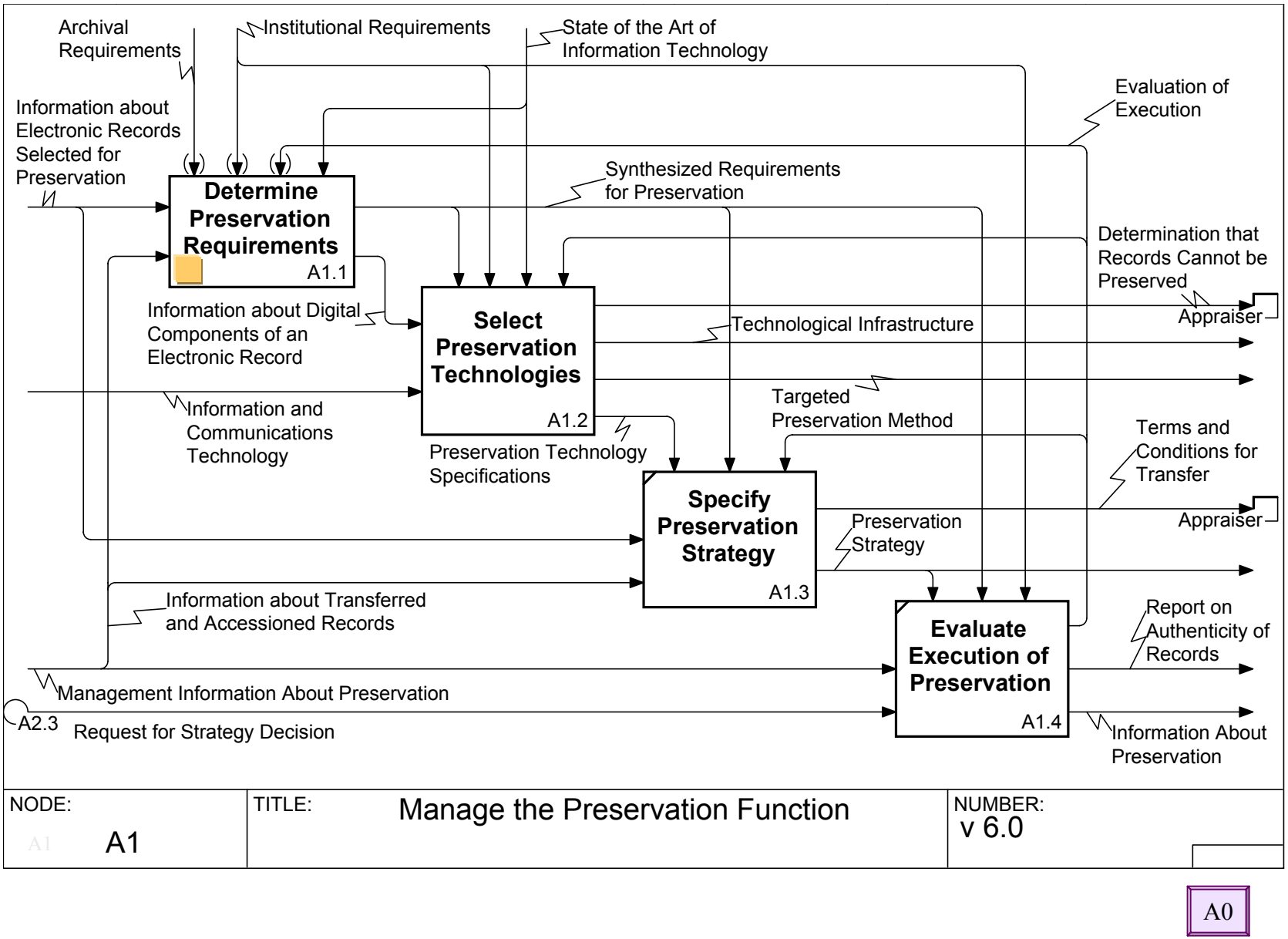
- Starting Point: the draft Open Archival Information System (OAIS) Reference Model
  - But: OAIS is not specific to records
- Purpose: to articulate what must be done, and what information and resources are needed, to preserve authentic electronic records.
  - But: requirements for authenticity were not available
- Viewpoint: the person responsible for (carrying out actions needed for) preserving electronic records.
- Scope: from the determination that records have long term value to the production of an authentic copy.
- Nature: Delineates a process, not a system or a workflow

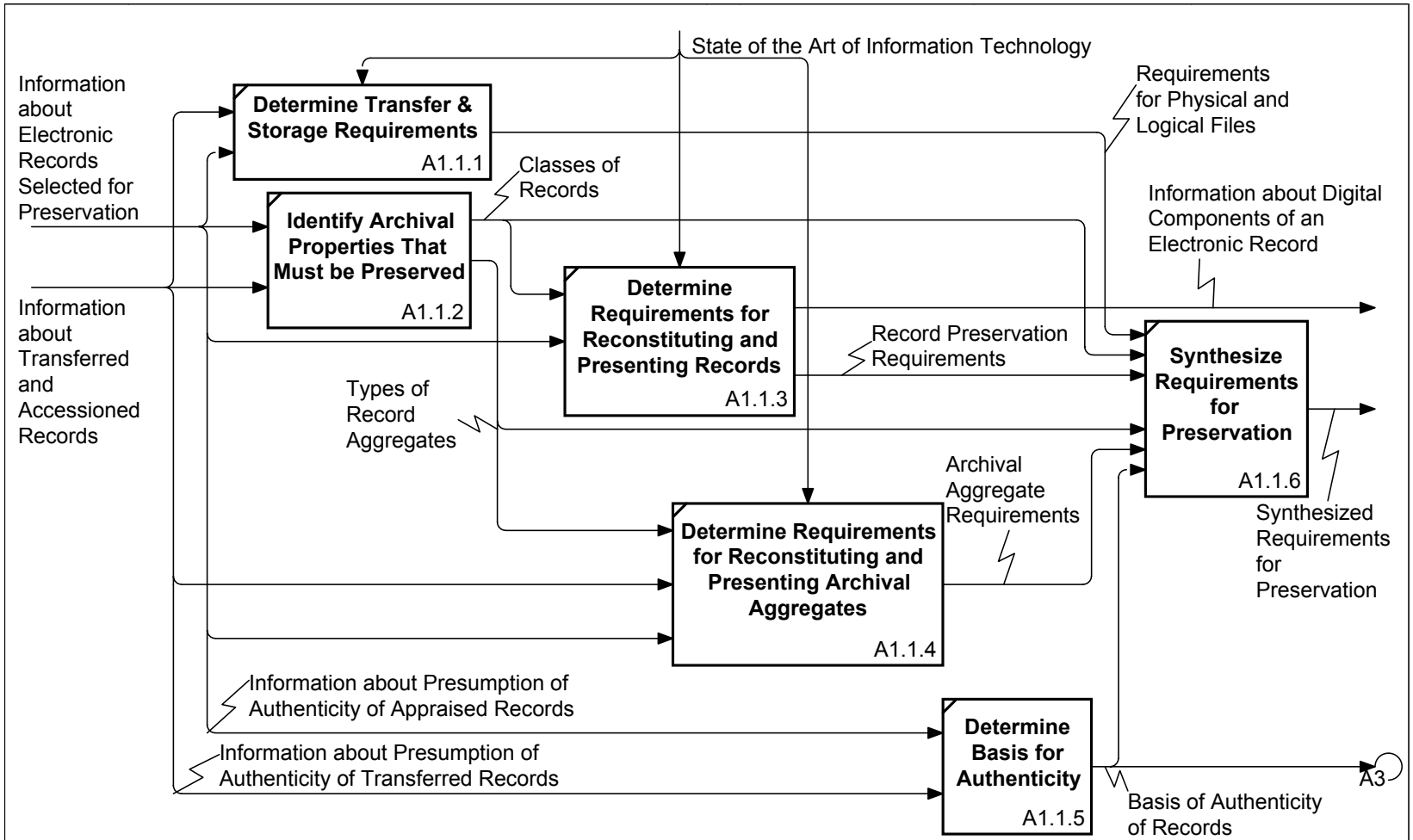


NODE:	TITLE:	NUMBER:
A-0	Preserve Electronic Records	v. 6.0



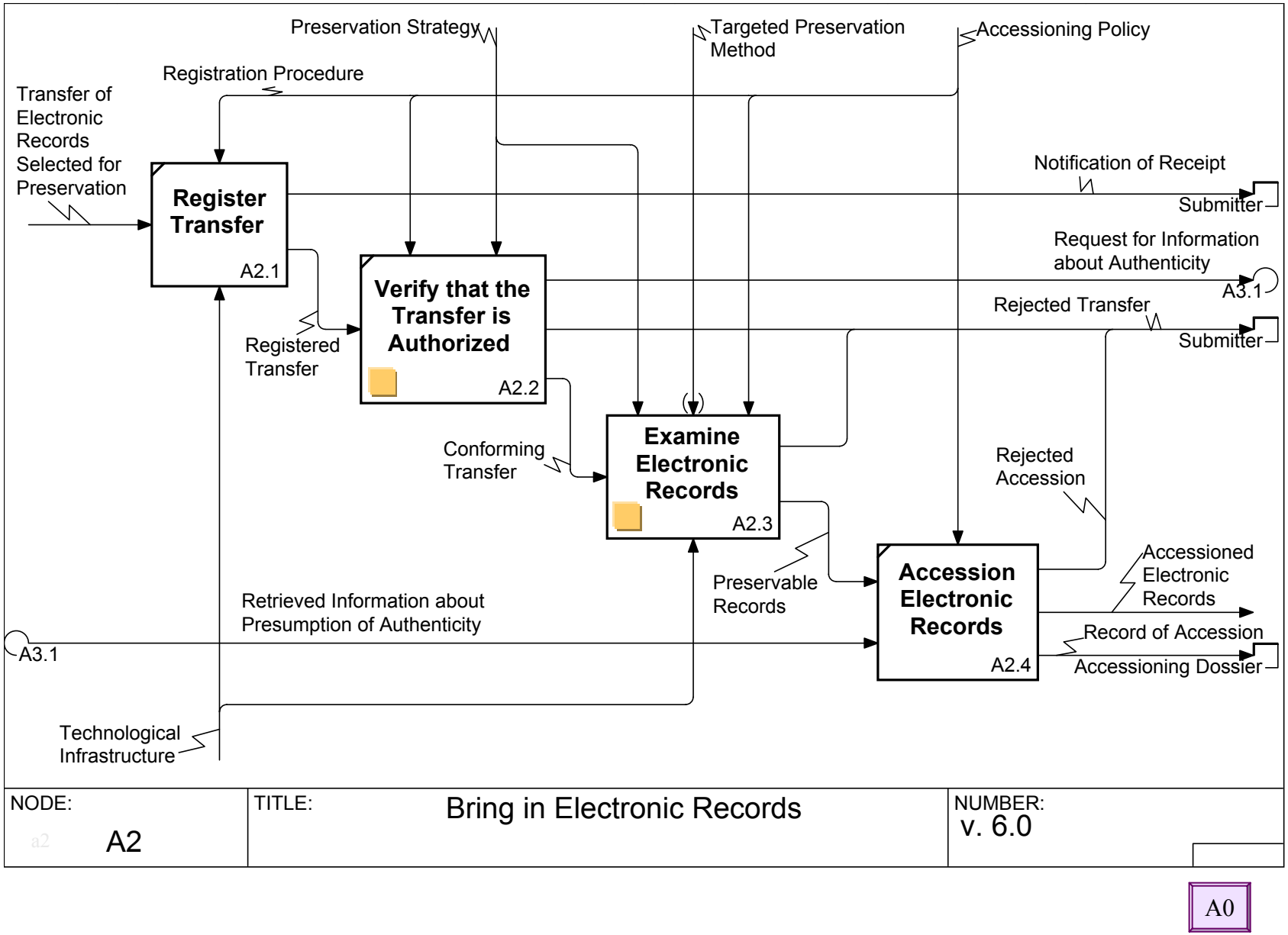
NODE:	TITLE:	NUMBER:
A0	Preserve Electronic Records	v 6.0

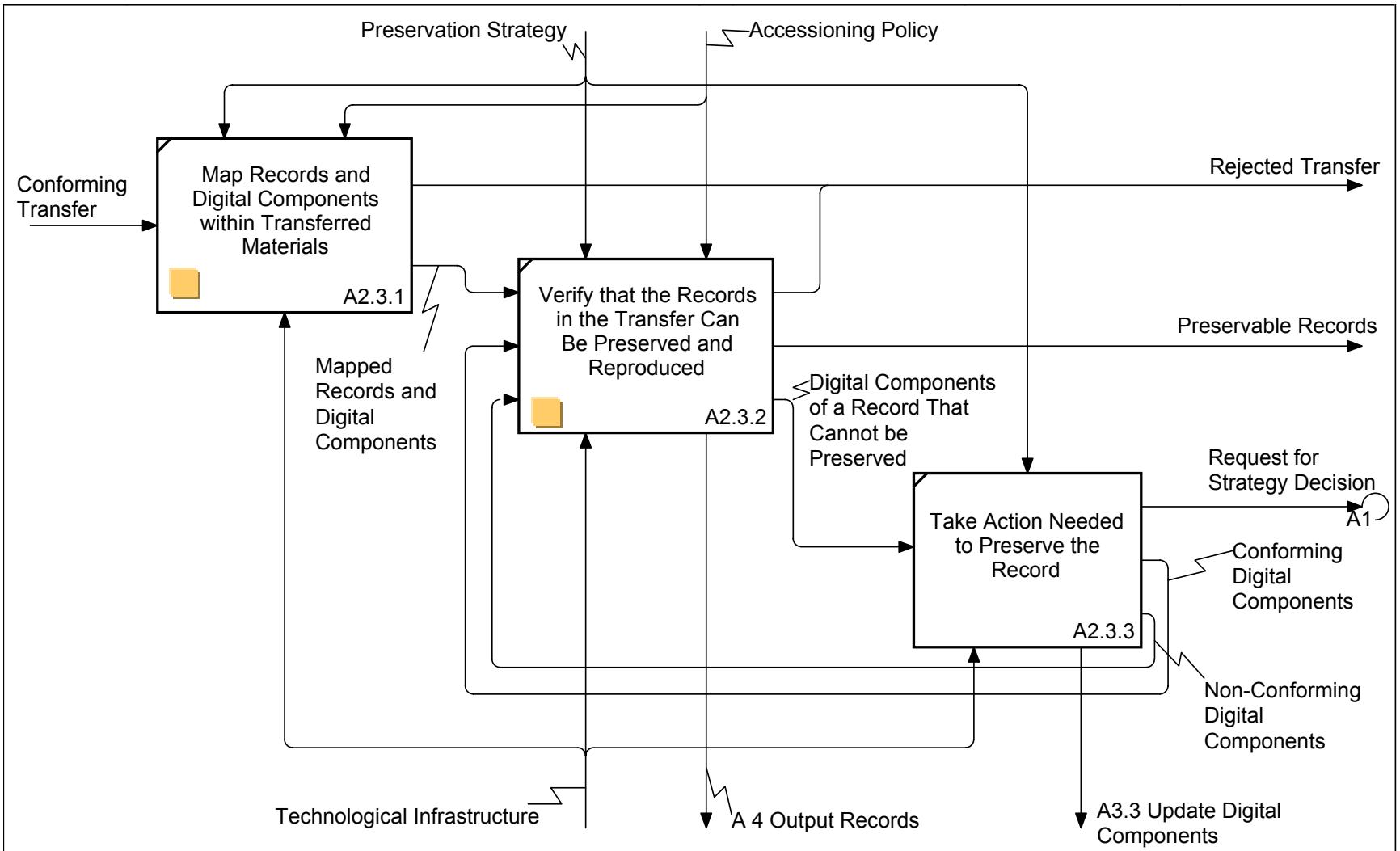

Archival Requirements
Institutional Requirements
State of the Art of Information Technology
Evaluation of Execution
Information about Electronic Records Selected for Preservation
Determine Preservation Requirements
A1.1
Synthesized Requirements for Preservation
Determination that Records Cannot be Preserved
Appraiser
Information about Digital Components of an Electronic Record
Select Preservation Technologies
A1.2
Technological Infrastructure
Information and Communications Technology
Targeted Preservation Method
Terms and Conditions for Transfer
Preservation Technology Specifications
Specify Preservation Strategy
A1.3
Appraiser
Preservation Strategy
Information about Transferred and Accessioned Records
Evaluate Execution of Preservation
A1.4
Report on Authenticity of Records
Management Information About Preservation
A2.3
Request for Strategy Decision
Information About Preservation
NODE:
A1
TITLE:
Manage the Preservation Function
NUMBER:
v 6.0
A0



NODE:	TITLE:	NUMBER:
A1.1	Determine Preservation Requirements	v. 6.0

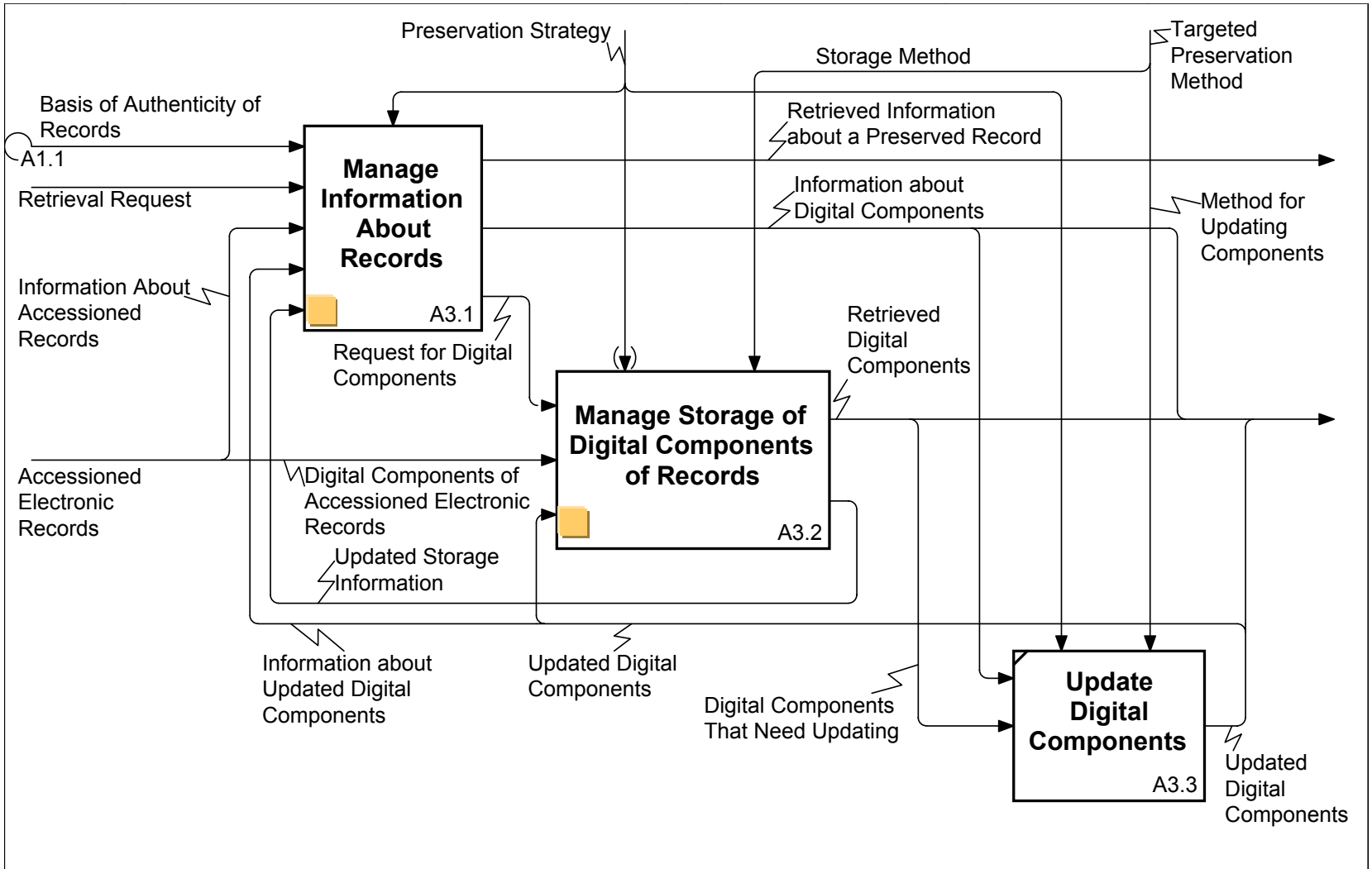
A1 A0


Preservation Strategy
Targeted Preservation Method
Accessioning Policy
Registration Procedure
Transfer of Electronic Records Selected for Preservation
Register Transfer
A2.1
Notification of Receipt
Submitter
Request for Information about Authenticity
A3.1
Verify that the Transfer is Authorized
A2.2
Registered Transfer
Rejected Transfer
Submitter
Conforming Transfer
Examine Electronic Records
A2.3
Rejected Accession
Accession Electronic Records
A2.4
Accessioned Electronic Records
Record of Accession
Accessioning Dossier
Preservable Records
Retrieved Information about Presumption of Authenticity
A3.1
Technological Infrastructure
NODE:
A2
TITLE:
Bring in Electronic Records
NUMBER:
v. 6.0
A0

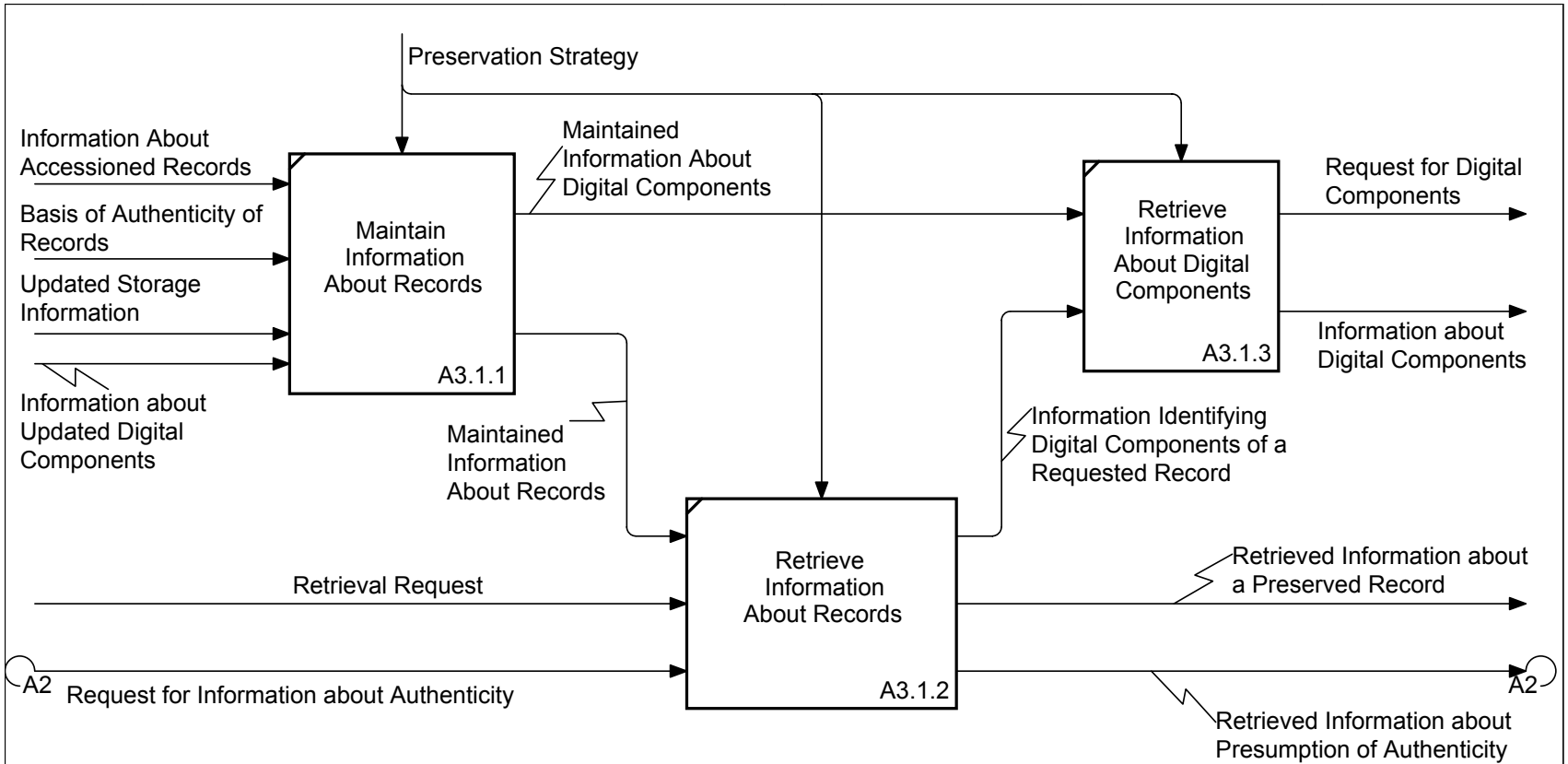


NODE:	TITLE:	NUMBER:
A2.3	Examine Electronic Records	v 6.0

A2 A0

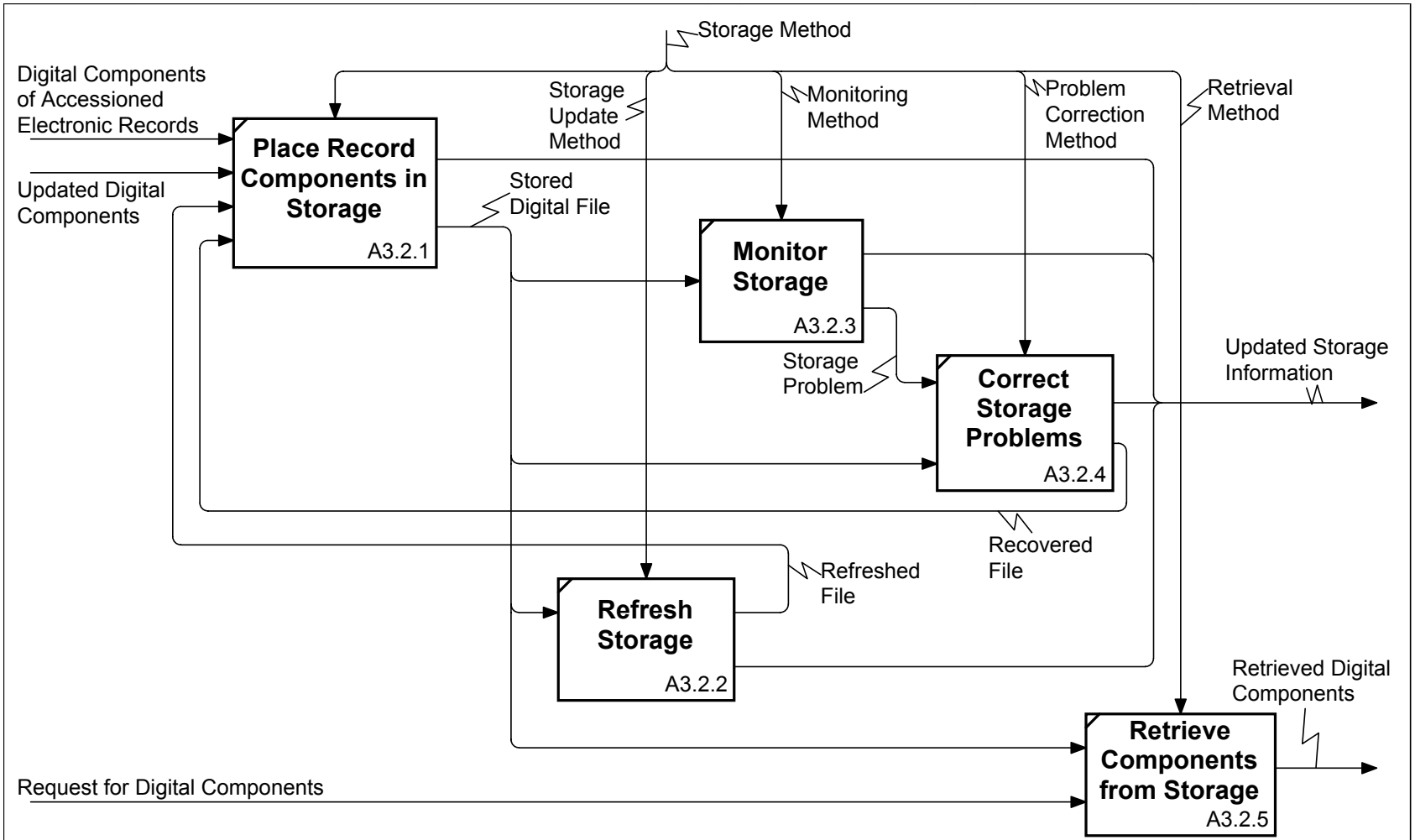


NODE:	TITLE:	NUMBER:
A3	Maintain Electronic Records	v 6.0



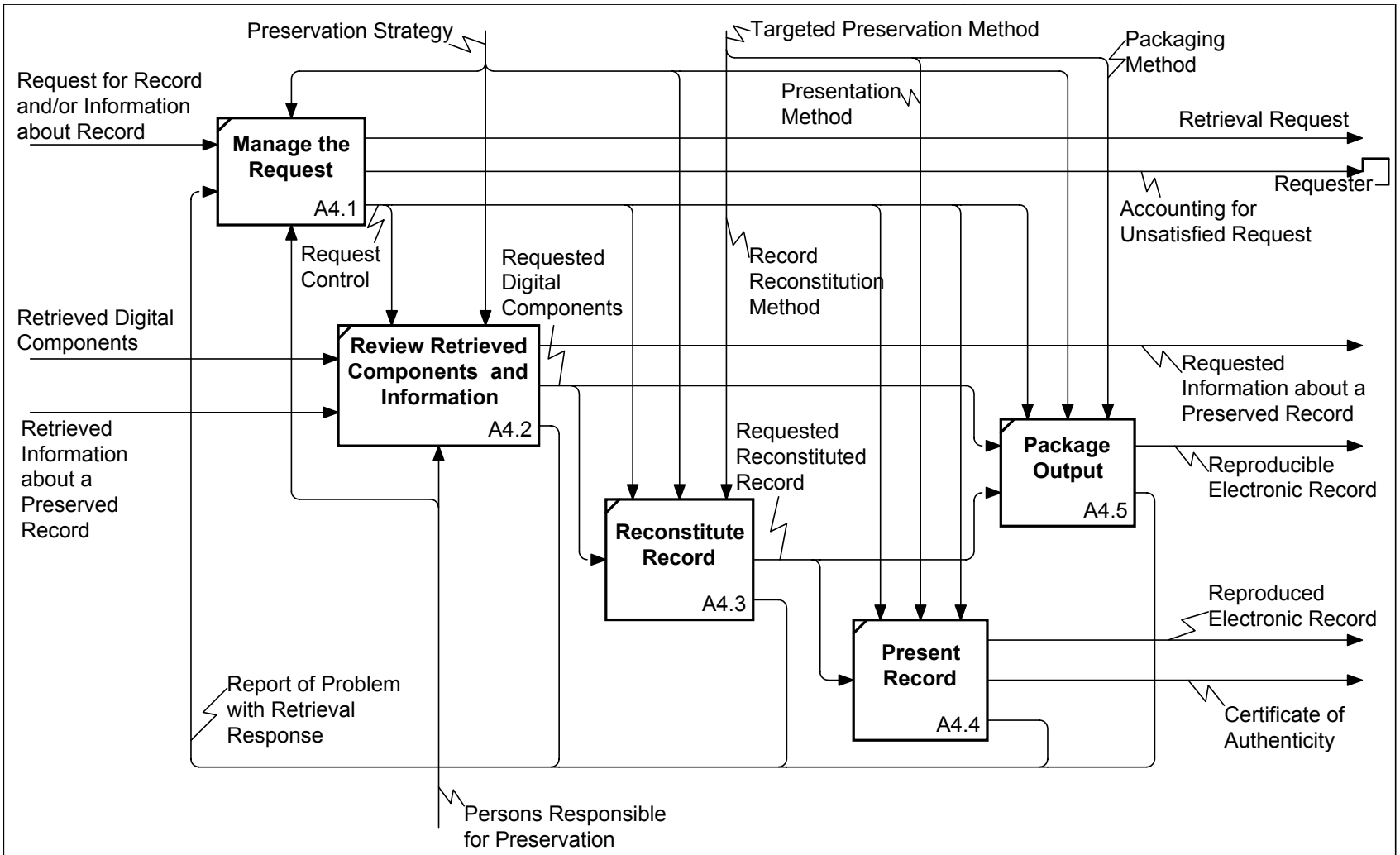
NODE:	TITLE:	NUMBER:
A3.1	Manage Information About Records	v 6.0

A3 A0



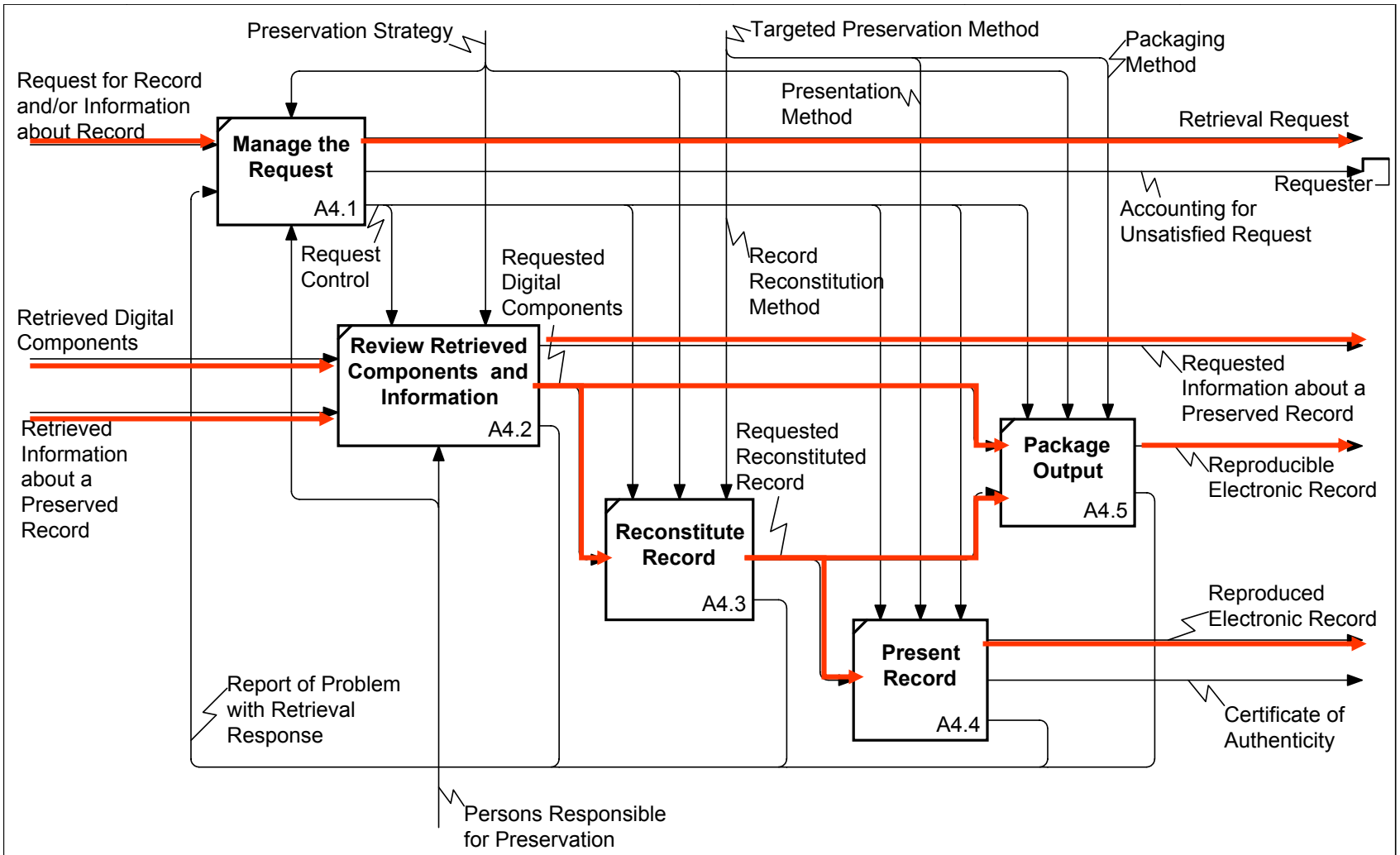
NODE:	TITLE:	NUMBER:
A3.2	Manage Storage of Digital Components of Records	v. 6.0

A3 A0



NODE:	TITLE:	NUMBER:
A4	Output Electronic Record	v 6.0

A0



NODE:	TITLE:	NUMBER:
A4	Output Electronic Record	v 6.0

A0

Verify that a transfer is authorized

- The Preservation Strategy (Control) ensures that terms and conditions for transfer are satisfied:
  - Who is authorized to send the records;
  - when should transfer occur;
  - what records should be transferred together;
  - what format(s) should the records be in;
  - what information should accompany the transfer.
- Check information accompanying a Registered Transfer (Input) to verify that these conditions are satisfied.
  - If so, the transfer is a Conforming Transfer (Output).
    - Request information about the basis for assuming that the creator maintained the records authentic (Output).
  - If not, reject the transfer, notifying the submitter (Output).

A2

Verify That Records Can Be Preserved and Reproduced

- Can each record be reconstituted from its digital components?
  - Does the data type of each component conform to the Preservation Strategy?
    - If not, should the transfer be rejected?
  - Do components in any data type need to be converted for preservation?
    - If so, refer for appropriate action.
- Can each record be output, in proper order with respect to related records?
  - If not, should the transfer be rejected?


ERA
Electronic Records Archives

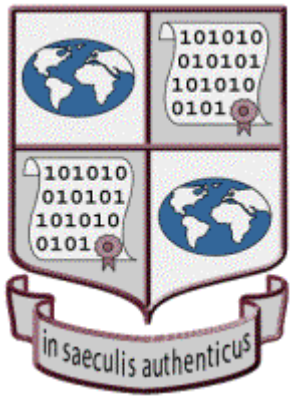
Mapping Records

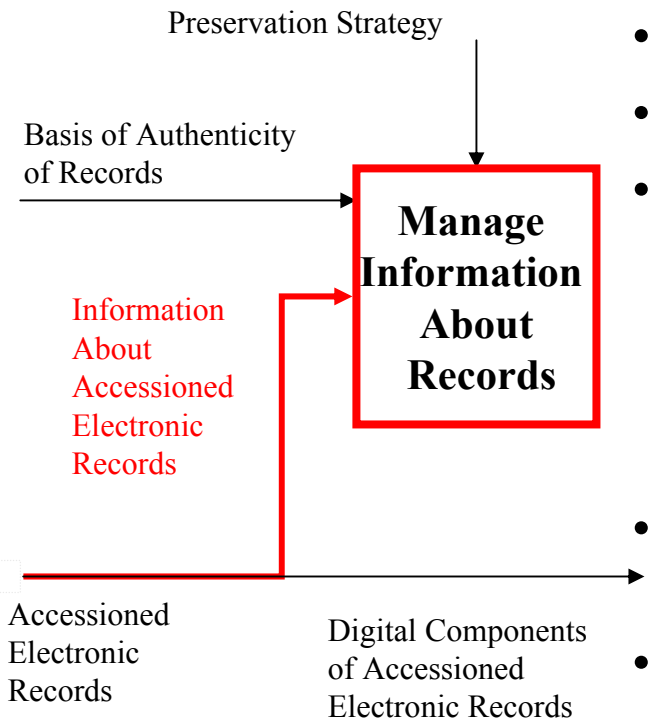
- What records and aggregates of records are reportedly included in a Transfer?
- What are the digital components of each record?
- Where are these components found in the physical file(s) transferred?
- Are all required components present?

  – *Does the transfer include any digital components that are not parts of records specified for the transfer?*

Example: Workers Compensation Board Case Folder System

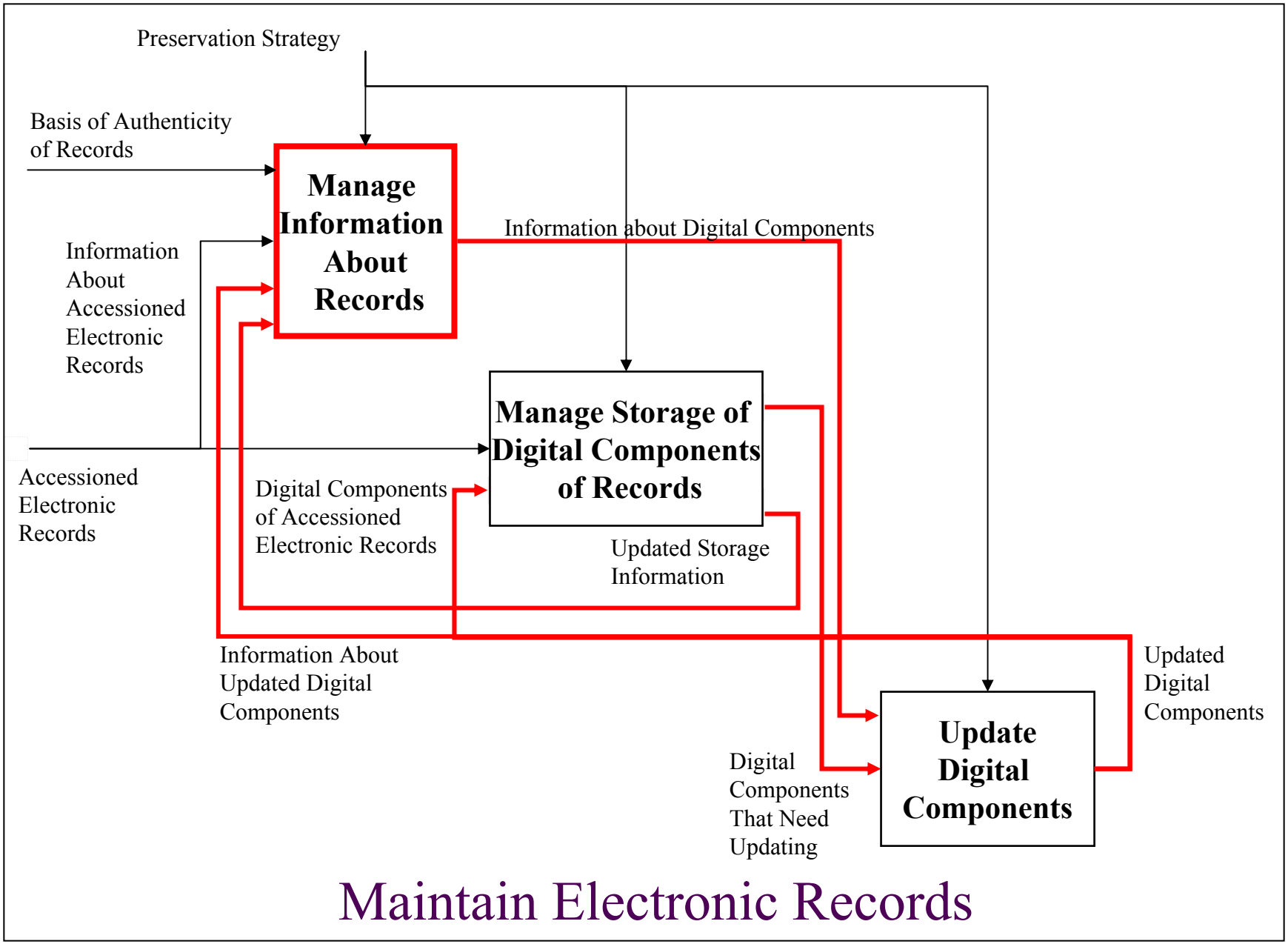
- Records
  - Aggregates: 1 series of case files
  - Records: 5 classes of documents
- Digital components
  - Each document stored as a multipage TIFF file
  - Relational database storing data about each document, documents in each case file, and case files in the series.
  - Metadata about the database


101010
010101
101010
0101
101010
010101
101010
0101
in saeculis authenticus

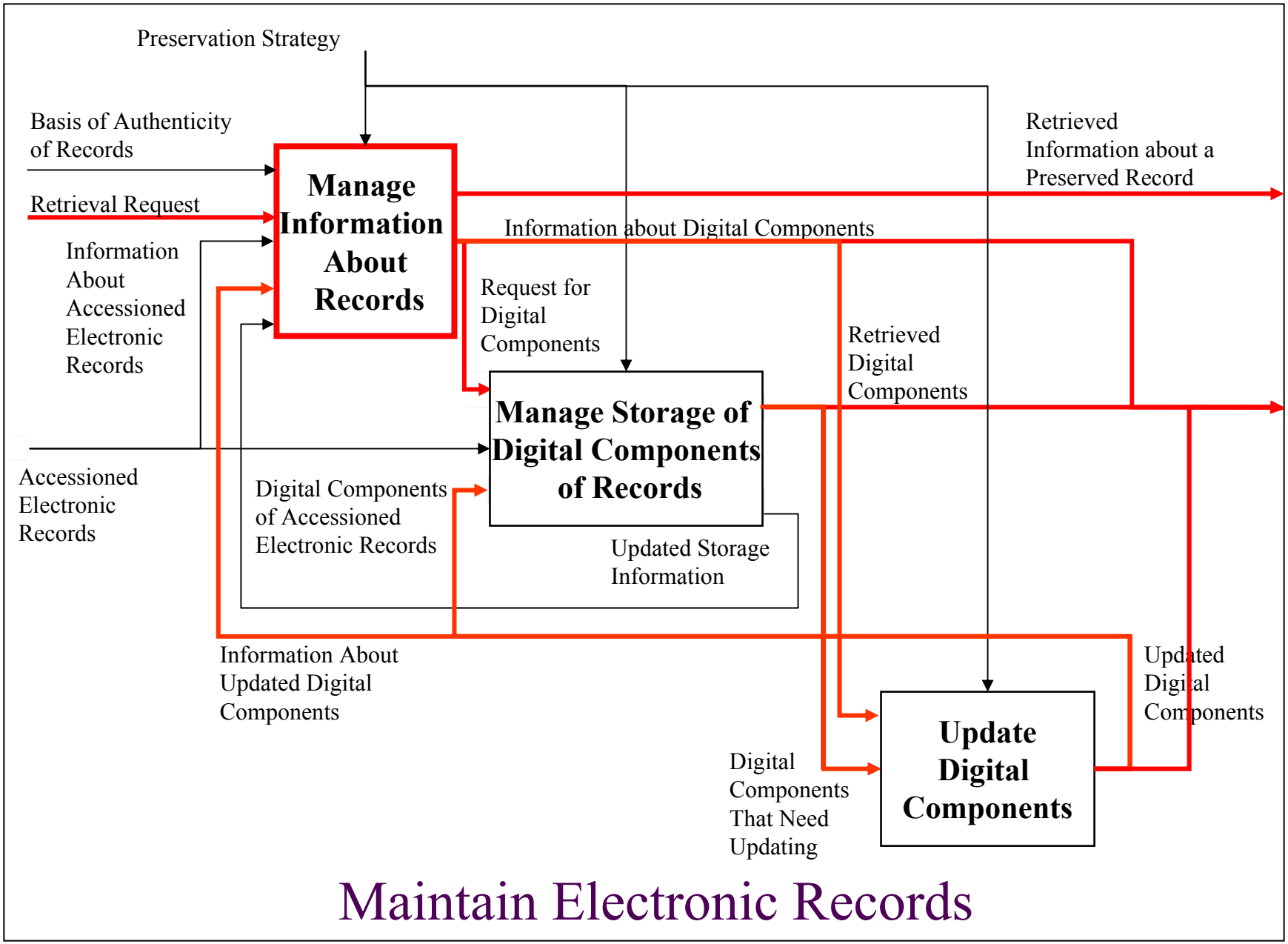


- Updated Basis of Authenticity
- Records: Case folders n through n+i
- Digital components:
  - Relational Database Tables
    - Metadata defining each table
    - Metadata defining relationship between tables
  - One TIFF file containing each document
- Storage: map of tables and TIFF files to physical files
- Preservation Information
  - Successful transfer
  - Successful updating of components, if any

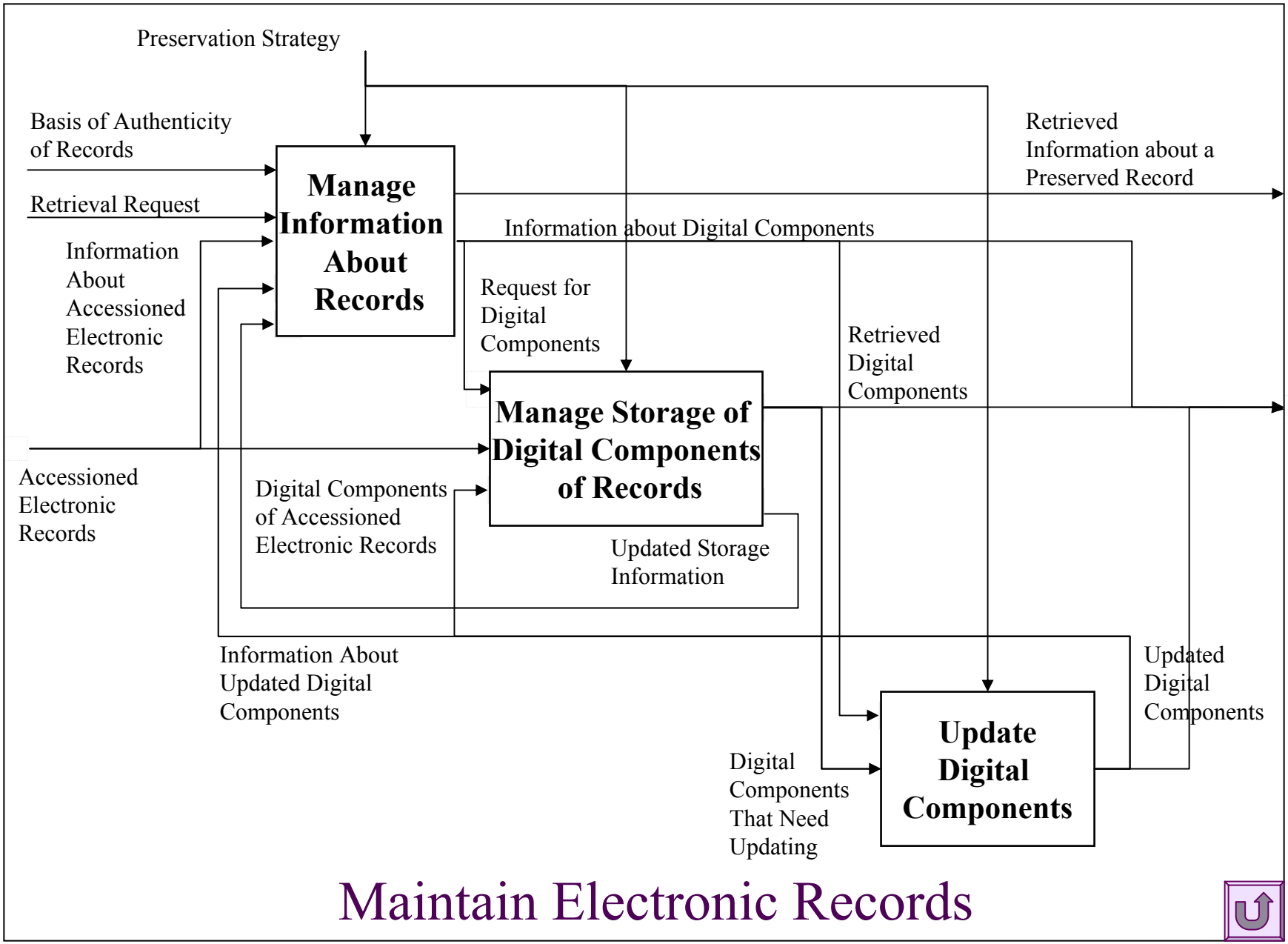
Maintain Electronic Records



Maintain Electronic Records



Maintain Electronic Records



Maintain Electronic Records