

Appendix D

Elsevier Science Technical Systems and Processes

Glossary for Standards

Distributed content from the ES warehouse in the Netherlands contains data that have been encapsulated or bundled in five different distribution formats that reflect the technological advancement of ES production and distribution process. The distribution datasets were once called Elsevier Electronic Subscriptions (EES), now obsolete, and were replaced in 1998 by Science Direct OnSite (SDOS): The version history is as follows:

PRECAP: Pre-computer aided production; placed into service in 1995

CAP: Computer Aided production; placed into service in 1997

EES V1.0

TIFF files containing scanned images
Raw ASCII text files, one for each page
SGML citation files
Dataset.toc file in EFFECT 4.0 specification

EES Version 1.1

Same as above except that the TIFF image page files were replaced by wrapped PDF files that contained an editorial item
Dataset.toc file in EFFECT 4.0 specification

EES Version 1.2

Same as EE version 1.1 but editorial items could be contained in wrapped PDF or true PDF format, i.e., converted from original Postscript file – highest resolution.
Dataset.toc file in EFFECT 4.0 specification

SDOS Version 2.0

PDF files containing an editorial item in wrapped or true format
Raw ASCII files containing an editorial item in wrapped or true format
SGML citation files containing bibliographic data for editorial items
Dataset.toc file in EFFECT 4.0 specification

SDOS Version 2.1

PDF files containing an editorial item in wrapped or true format
Raw ASCII files containing an editorial item in wrapped or true PDF format
SGML citation files containing bibliographic data for editorial items and article references in structured format
Dataset.toc file in EFFECT 4.0 specification

SDOS Version 3.0

PDF files containing a publication item in wrapped or true format
Raw ASCII files containing a publication item in wrapped or true format
Full article SGML files for publication items, artwork files in Web-enabled graphical formats
Dataset.toc file in EFFECT 4.1 specification

Data Components Found in EES and SDOS Datasets

Page Images

Black and white
TIFF 5.0 standard
Scanned at 300 dpi
Maximum scan is European A4, i.e., 210x297mm²
Compression ITU T.6, aka CCITT Fax group 4, for an average page 8%
Compression is achieved, i.e., 1M \pm 80Kbytes
White background and black characters

Raw Text Files

Each page image has a corresponding raw ASCII file
Produced from OCR procedures
No keyboarding/editing/spell-checking is performed on them
Contain only ASCII characters 32-126
Provided as a basis for searchable indexes – not for end users

SGML Files

Text of editorial items

SGML files are encoded in plain ASCII

SGML files have two extension attributes: ".sgc" and ".sgm"

Former means SGML data for heading information and the latter means full SGML content

Note: SDOS2.1 contains only ".sgc" files

Other Files

Pertains to distribution of content. Supplier and receiver agree that files with these other formats for content can be packaged in SDOS 2.1 datasets.

Adobe Acrobat Portable Document Format (PDF) Item/Page basis. Item based files contain a one-to-one ratio of one PDF file for one issue article.

Page-based PDF files contain pages that are not part of a clearly identified item/article such as front and back covers, advertisements etc.

Together item-based and page based PDF files can be used to reconstruct the entire paper journal in electronic format.

True/Distilled: original typesetter Postscript files

- no paper scanning steps

- same quality as final paper journal issue

Wrapped: image scanning on the paper journal issue

- TIFF images - fax group 4 encapsulated in PDF code

- lesser quality than distilled

Encapsulated PostScript (EPS)

Joint Photographer Expert Group (JPEG) encoded files

Hypertext Markup Language files

CompuServe Graphics Interchange Format (GIF) compressed files

TEX encoded files

CHECKMD5.FIL: Checksum facility to ensure the validity or integrity of the data distributed to the Client.

EFFECT- DATASET.TOC FILE:

Contains all cross-indexing reference data needed to load into an application or database. See EFFECT document for general rules of this file.

DATASET.TOC is split up into records that are broken into four major divisions.

- _t0 → all data on the complete dataset
- _t1 → all data on a specific journal title
- _t2 → all data on a specific journal issue of title _t1
- _t3 → the first editorial item within the issue
- _t3 → the second editorial item within the issue
- _t2 → the second journal issue
- _t3 → the first editorial item within the issue
- _t1 → another journal title

Diagram I - Electronic Warehouse Hierarchy

ES Group
Chart: 030301 Prod Prim Ams.vsd
Updated: 1 February 2001

Production Primary Publications Amsterdam

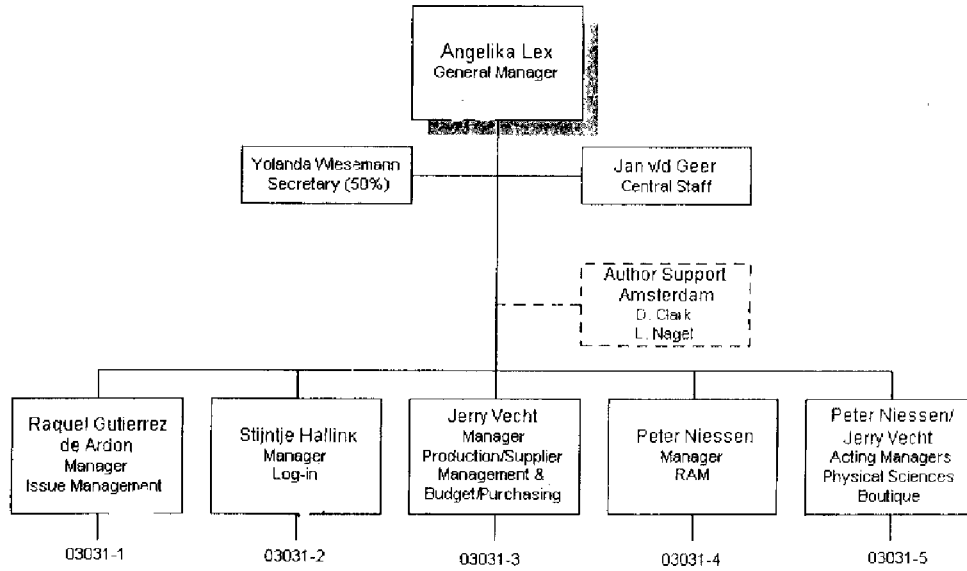


Diagram II - Computer Aided Workflow

The CAP Workflow

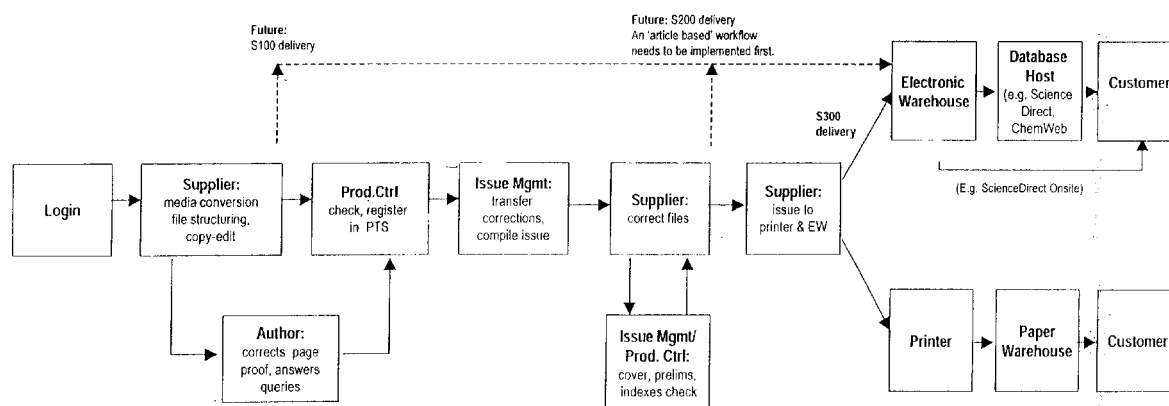


Diagram III - E-Workflow Overview

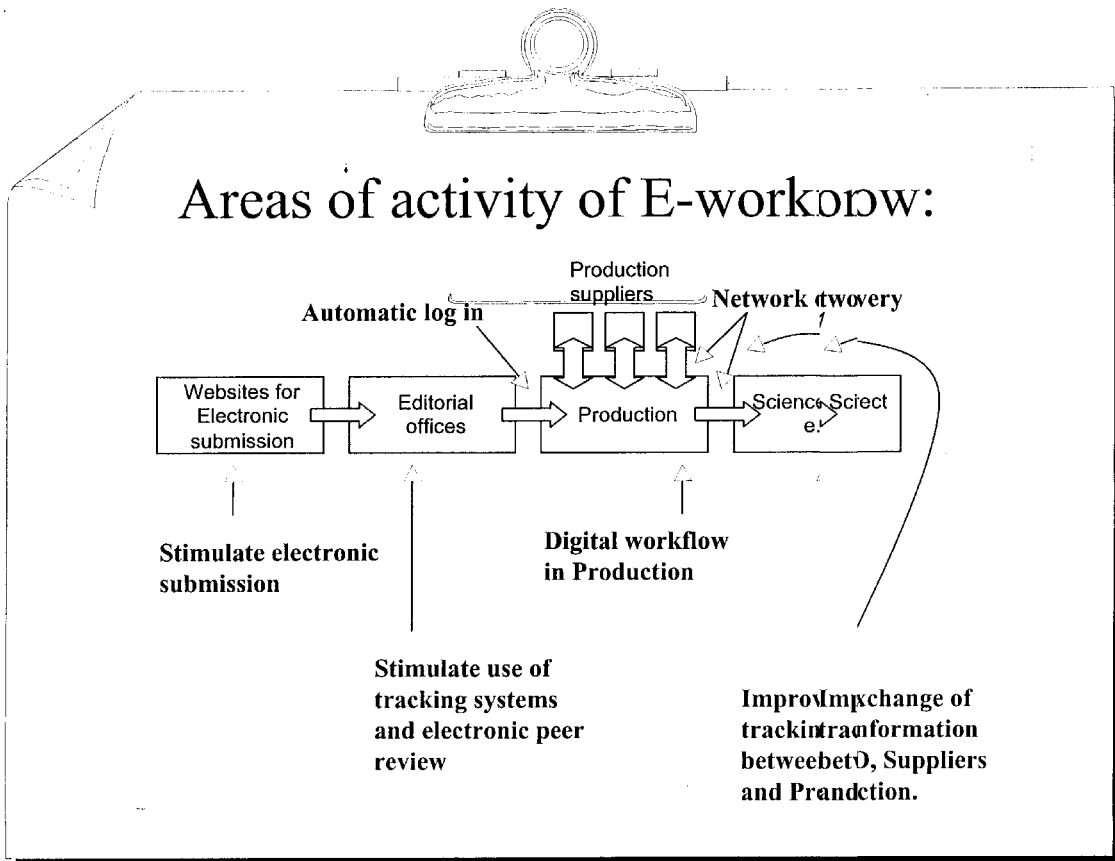
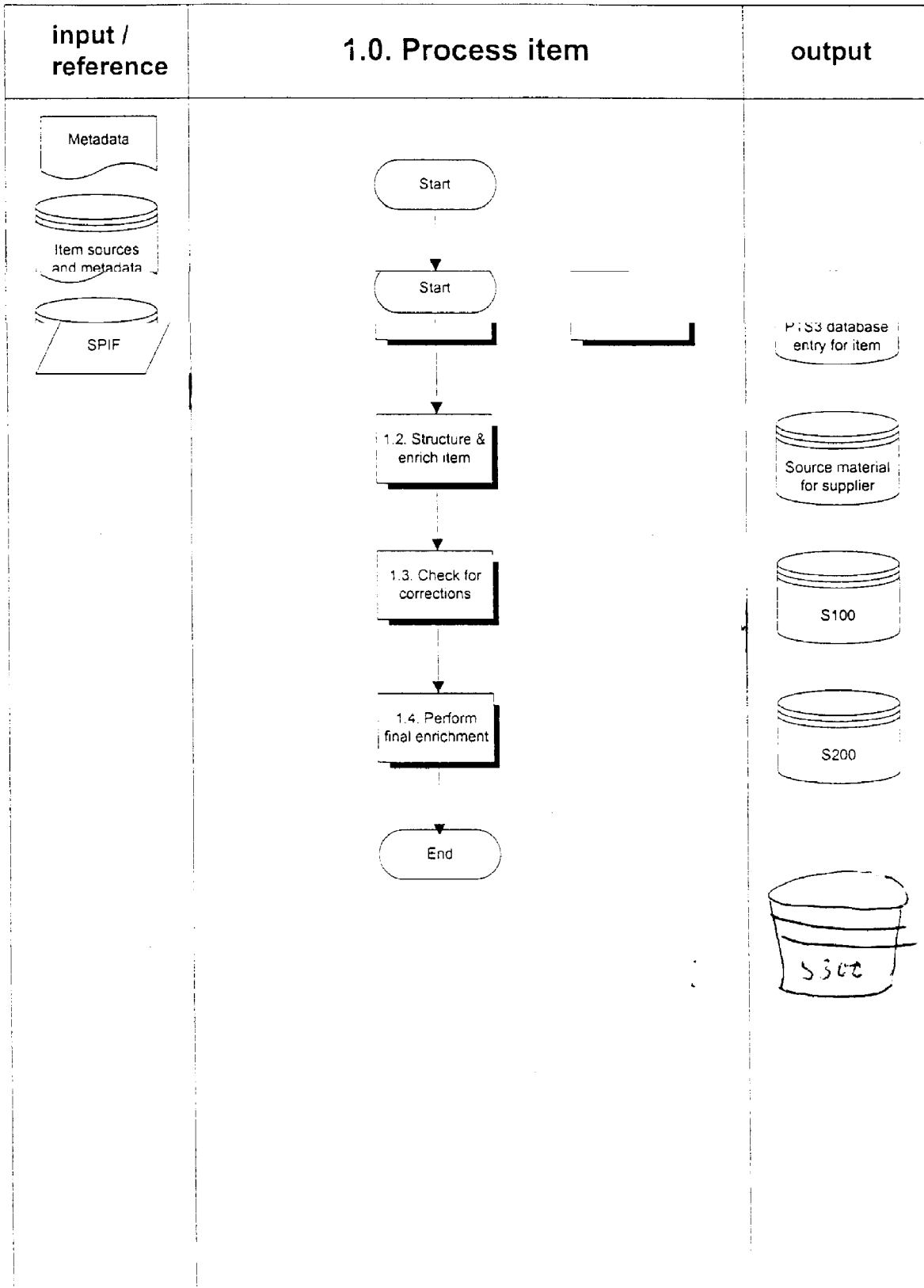
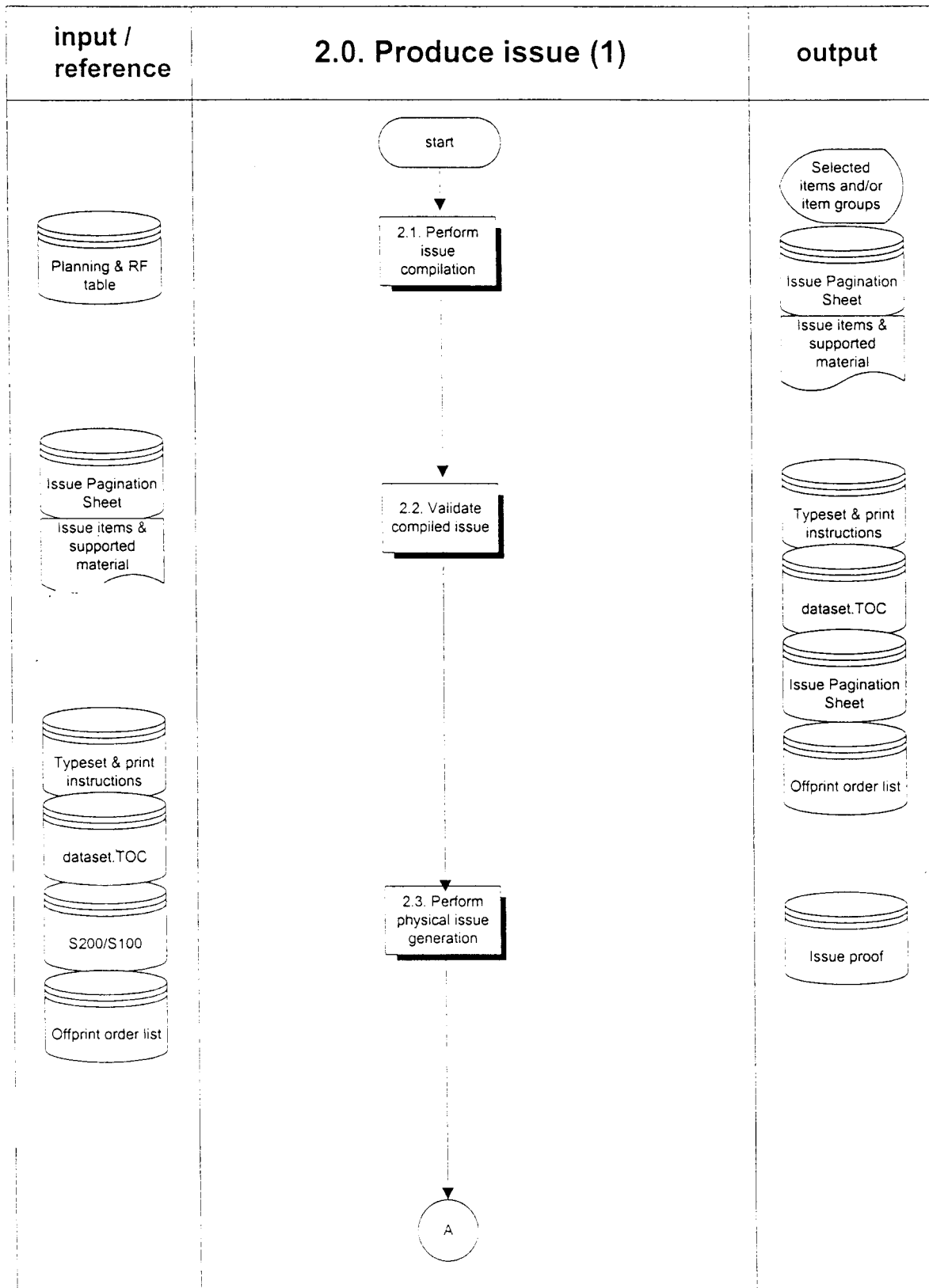
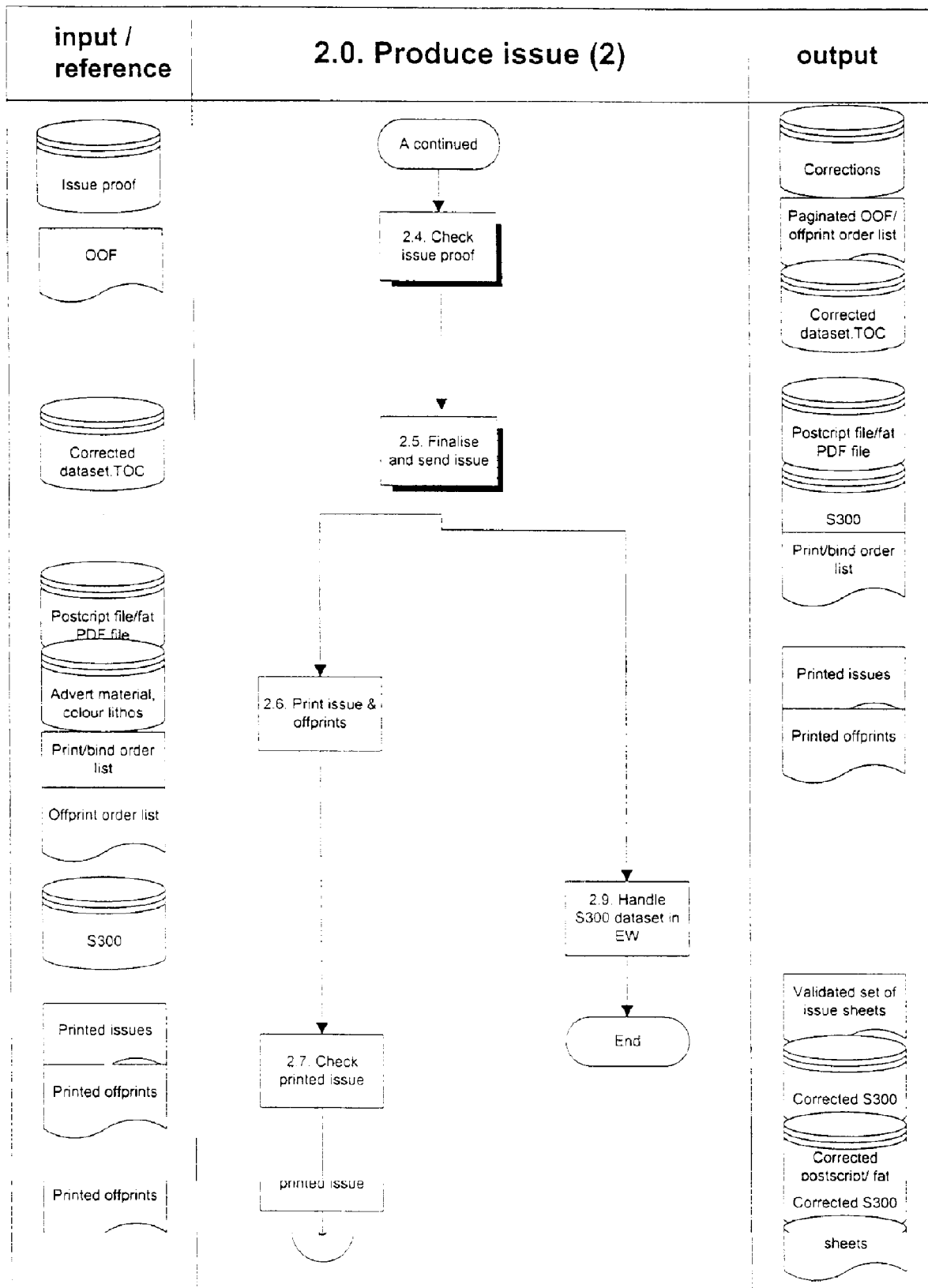


Diagram IV - E-Workflow Flow Charts







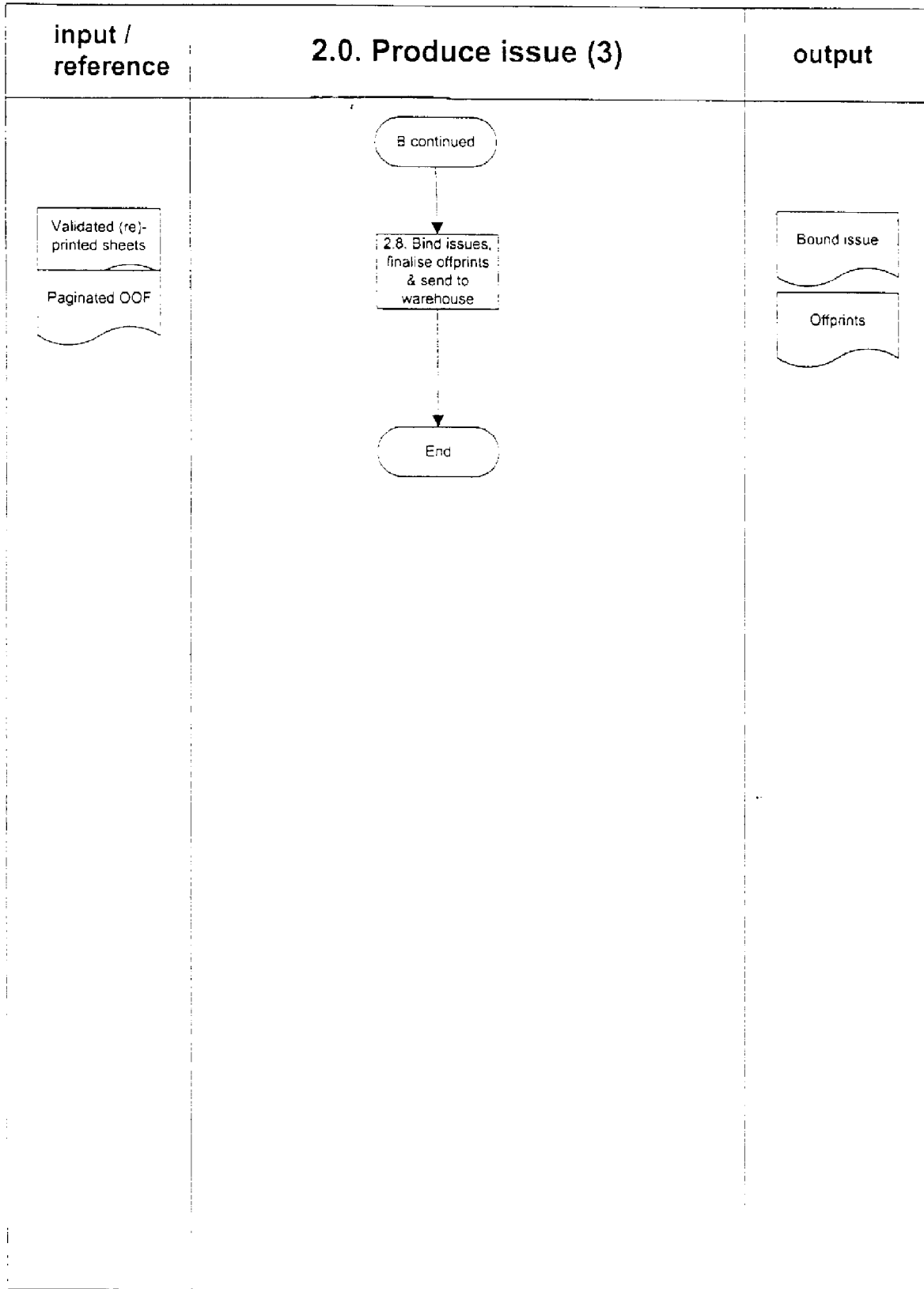


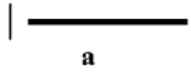
Diagram V- Science Direct Trigger or Tracking Codes

2 Possible codes

1. New
2. Acquired from
3. Continued as
4. Formerly known as
5. Split into
6. Formerly part of
7. See also
8. Formerly included in
9. Merged into
10. Merger of
11. Incorporated into
12. Incorporating
13. Transferred to
14. Discontinued

1 Journal events

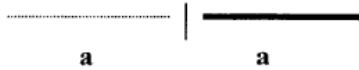
1.1 New journal



Required components:

ISSN a	Jtitle a	Year	First Volume	First Issue	New				
-----------	-------------	------	-----------------	----------------	------------	--	--	--	--

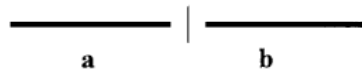
1.2 Acquired journal



Required components:

ISSN a	Jtitle a	Year	First Volume	First Issue	Acquired from			Former Owner	1 st Year of Publi- cation
-----------	-------------	------	-----------------	----------------	--------------------------	--	--	-----------------	--

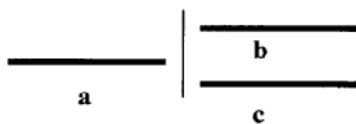
1.3 Journal rename



Required components:

ISSN a	Jtitle a	Year	Last Volume	Last Issue	Continued as	ISSN b	Jtitle b		
ISSN b	Jtitle b	Year	First Volume	First Issue	Formerly known as	ISSN a	Jtitle a		

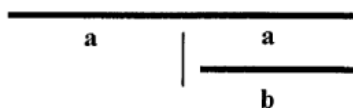
1.4 Journal split



Required components:

ISSN a	Jtitle a	Year	Last Volume	Last Issue	Split into	ISSN b	Jtitle b		
ISSN a	Jtitle a	Year	Last Volume	Last Issue	Split into	ISSN c	Jtitle c		
ISSN b	Jtitle b	Year	First Volume	First Issue	Formerly part of	ISSN a	Jtitle a		
ISSN c	Jtitle c	Year	First Volume	First Issue	Formerly part of	ISSN a	Jtitle a		

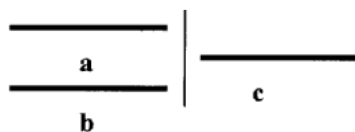
1.5 Journal spin-off



Required components:

ISSN a	Jtitle a	Year	Last Volume	Last Issue	See also	ISSN b	Jtitle b		
ISSN b	Jtitle b	Year	First Volume	First Issue	Formerly included in	ISSN a	Jtitle a		

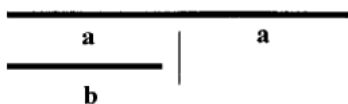
1.6 Journal merger



Required components:

ISSN a	Jtitle a	Year	Last Volume	Last Issue	Merged into	ISSN c	Jtitle c		
ISSN b	Jtitle b	Year	Last Volume	Last Issue	Merged into	ISSN c	Jtitle c		
ISSN c	Jtitle c	Year	First Volume	First Issue	Merger of	ISSN a	Jtitle a		
ISSN c	Jtitle c	Year	First Volume	First Issue	Merger of	ISSN b	Jtitle b		

1.7 Journal incorporation



Required components:

ISSN a	Jtitle a	Year	First Volume	First Issue	Incorporating	ISSN b	Jtitle b		
ISSN b	Jtitle b	Year	Last Volume	Last Issue	Incorporated into	ISSN a	Jtitle a		

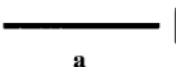
1.8 Transferred journal



Required components:

ISSN a	Jtitle a	Year	Last Volume	Last Issue	Transferred to			New Owner	
-----------	-------------	------	----------------	---------------	----------------	--	--	--------------	--

1.9 Discontinued journal



Required components:

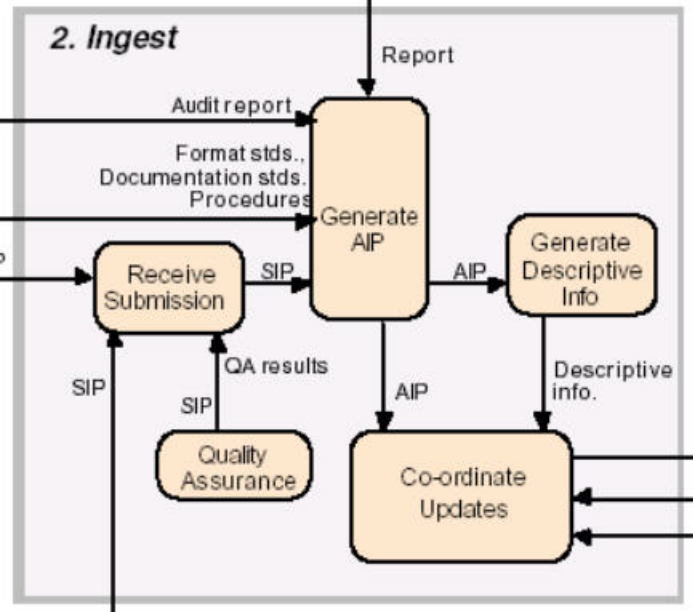
ISSN a	Jtitle a	Year	Last Volume	Last Issue	Discontinued				
-----------	-------------	------	----------------	---------------	--------------	--	--	--	--

Appendix E

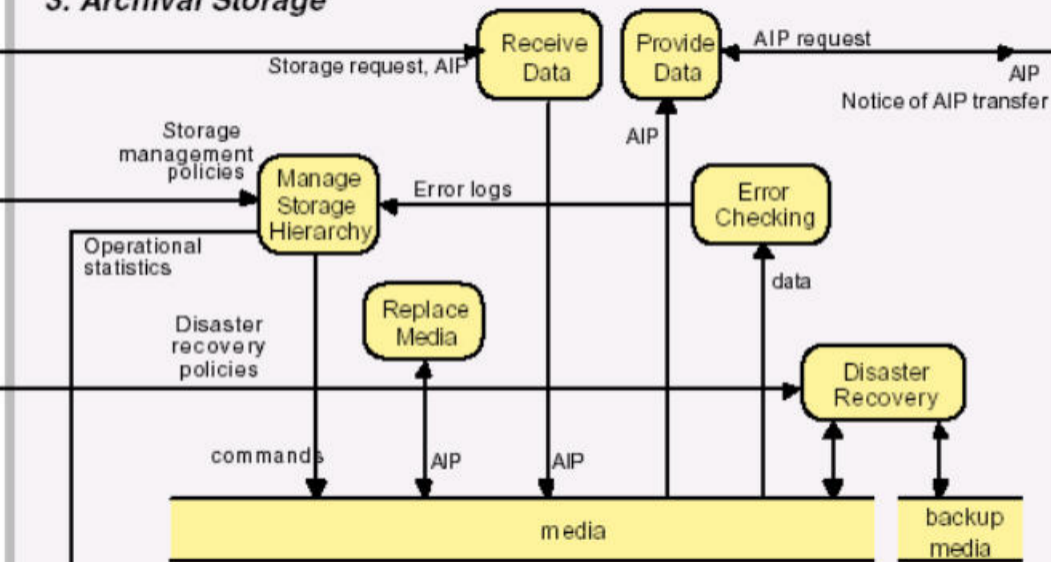
OAIS Diagrams and Spreadsheet

The following OAIS diagrams provide details about the processes and functions that make up the OAIS reference model. During the planning year, much of the project work was concerned with the ingestion and preservation planning processes. The corresponding spreadsheet entitled *OAIS Model: Areas of Investigation in Mellon Planning Year* identifies specific processes and functions that were investigated in the first year of project work. In phase two, the YEA team will continue to work in these areas but will also begin research on areas that were not investigated in phase one. These are also identified, in the spreadsheet.

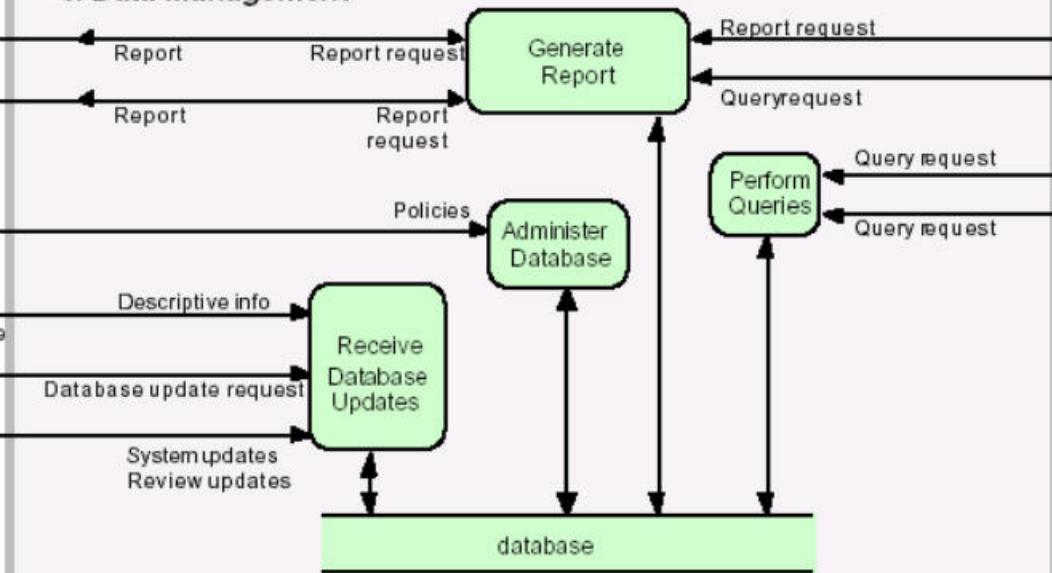
2. Ingest



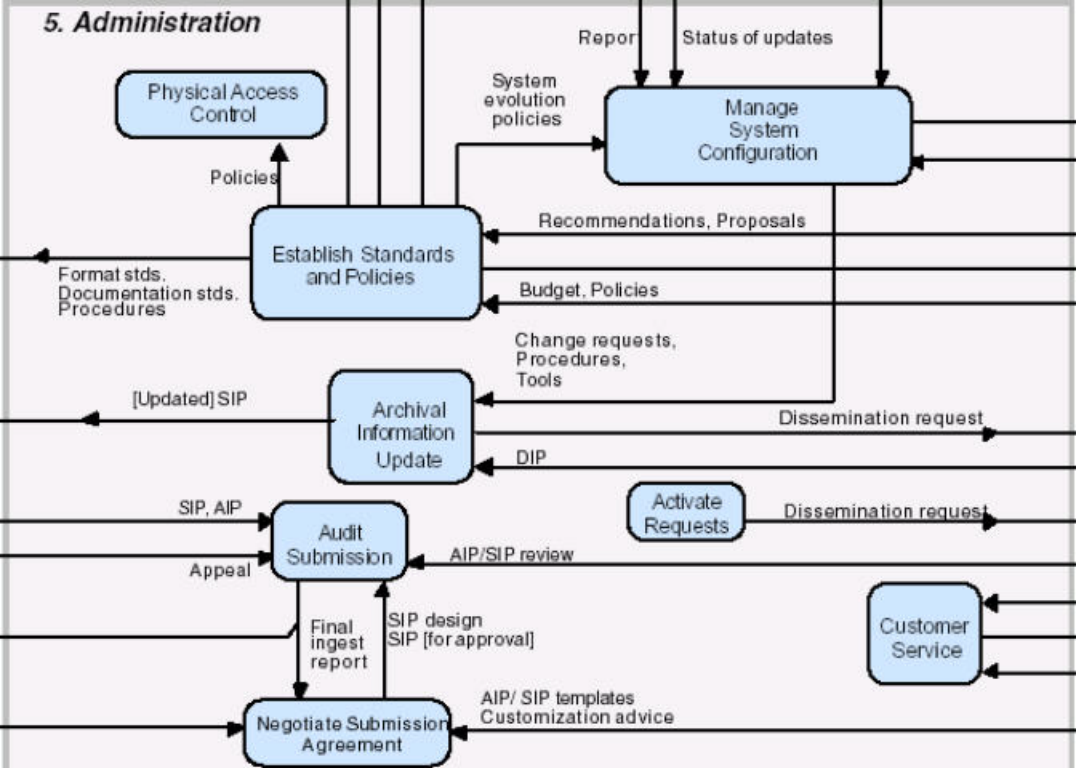
3. Archival Storage



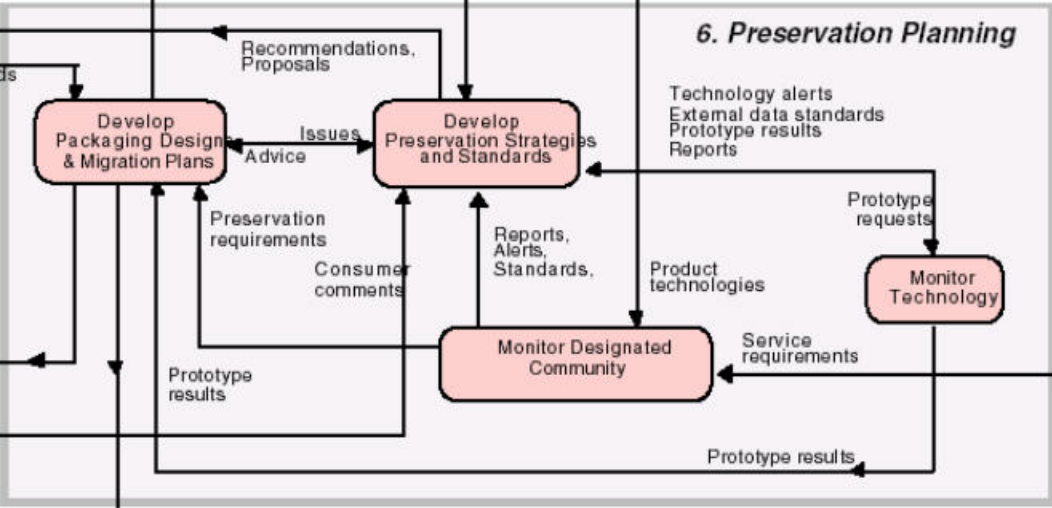
4. Data Management



5. Administration



6. Preservation Planning



OAIS Model: Areas of Investigation in Mellon Planning Year

OAIS Processes

OAIS Functions	Ingest Process	Archival Storage Process	Data Management Process	Administration Process	Preservation Planning Process	Access Process
Receive Submission	x,a,					
Generate AIP	x,a,c,					
Quality Assurance	x,b					
Generate Descriptive Info	x,b					
Coordinate Updates	x,b					
Receive data		x,a,c				
Managage Storage		x,a,c				
Replicate Media		x,b				
Provide Data		x,a,c				
Error Checking		x,a,c				
Disaster Recovery		x,b				
Backup		x,b				
Create Database			x,a,c			
Administer Database			x,b			
Receive database updates			x,b			
Generate reports			x,b			
Perform queries			x,a,c			
Negotiate Submission Agreements				x,a,c		
Physical Access Control				x,b		
Manage System Configuration				x,b		
Establish Standards and Policies				x,a,c		
Archival Information Update				x,b		
Audit Information Update				x,b		
Audit Submissions				x,b		
Customer Service				x,b		
Design Packaging and Mig. Plans					x,b	
Dev. Preser. Strategies					x,b	
Monitor Designated Community					x,a	
Monitor Technology					x,a,c	
Generate Dip						x,a,c
Co-Ordinate Access Activities						x,b
Deliver Response						x,a,c

Table Key:

- (1) x: Identifies an OAIS functions by process
- (2) A: Identifies an area of technical investigation in the planning year
- (3) B: Identifies areas not investigated in planning year
- (4) c: Identify prototype components that were built or process created in planning year

Appendix F

List of Site Visits During Planning Year

Date	Organization	Location	Purpose
March 26-30 2001	Elsevier Science National Library of the Netherlands	Amsterdam Den Haag	Fact finding trip to learn about the production of electronic journals by ES and to learn about the digital archive work being done at the National Library of the Netherlands.
September 6-11 2001	British Library National Library of the Netherlands	London Den Haag	Validation of OAIS and OAI models to build prototype archives; learn about best practices from sites that have ongoing archival programs.
May 7 2001	JP Morgan Chase	Yale University New Haven, CT	Fact finding visit to learn about potential economic benefits of outsourcing the storage component of an archive.
September 27 2001	WGBH	Boston, MA	Fact finding visit to discuss archival models such as OAIS, OAI and METS.
October 11-12 2001	Elsevier Science	Amsterdam	Fact finding trip to learn more about potential content beyond traditional journals, the production process, and metadata population.

Appendix G

Possible Structure for the Yale Digital Library, Illustrating Library Functions: Key to Diagram

Desirable Characteristics of a Digital Library Infrastructure

- Integration of system components
- Consolidation or aggregation of proliferating stand-alone databases
- Integration of a wide variety of digital objects and metadata schemas
- Integration of search interfaces and delivery mechanisms
- Flexible output: general, specialized, and personalized interfaces
- Interoperability with external systems and institutions
- Scalability
- Versatility
- Sophisticated management tools
- Direct focus on teaching and research needs

The diagram illustrates selected existing systems in the Yale Library and explores several future directions, with a focus on digital preservation. At the heart of the diagram is a new preservation archive for digital objects and associated metadata based on the Open Archival Information System model. The public interfaces on the right interact directly with this preservation archive. Those on the left rely upon completely independent systems where metadata and digital objects are stored separately from the archive. Content sources in the second column feed these systems in various ways outlined below.

Journal Publisher (Elsevier)

- Public access through full-featured online system maintained by vendor
- Formal partnership between Yale and vendor for archiving journal content
- Limited access to archive through OAI interface (Open Archives Initiative)

Digitized Content (Visual Resources, Beinecke, Digital Conversion Facility, Divinity, etc.)

Public interface supplies sophisticated visual environment for teaching and study

- Insight system houses derived images (JPEGs and SIDs) and public metadata
- Archive houses original TIFF images and enhanced metadata
- Archive used only for image recovery or migration to new delivery platform

Electronic Yale University Records

- University records preserved for legal and historical purposes, low-use material
- Content sent directly to archive; no duplication of data in separate system
- Public interface retrieves digital objects directly from archive

Born-Digital Acquisitions

- Content is imported or directly input into new public repository

- Potential home for digital scholarship resulting from collaborative research projects

- Archival copies are transmitted from there to the archive

Finding Aids

- Finding aids distributed both to public service system and to archive

Preservation Reformatting (Digital Conversion)

- Digitized content sent directly to archive

- Hard-copy may be produced from digital version for public use

- Access to digital copy through custom application fed from archive

- Digital copy and original artifact may appear in national registry

Online Catalog

- Cataloging data resides only in LMS (NOTIS or Endeavor Voyager)

- Integration achieved through MetaLib portal and lateral SFX links

