InterPARES 3 Project

International Research on Permanent Authentic Records in Electronic Systems

TEAM Canada

Digital Weft

InterPARES as common thread in Vancouver's Digital Preservation Programme

CITRA Oslo
Sept 15, 2010

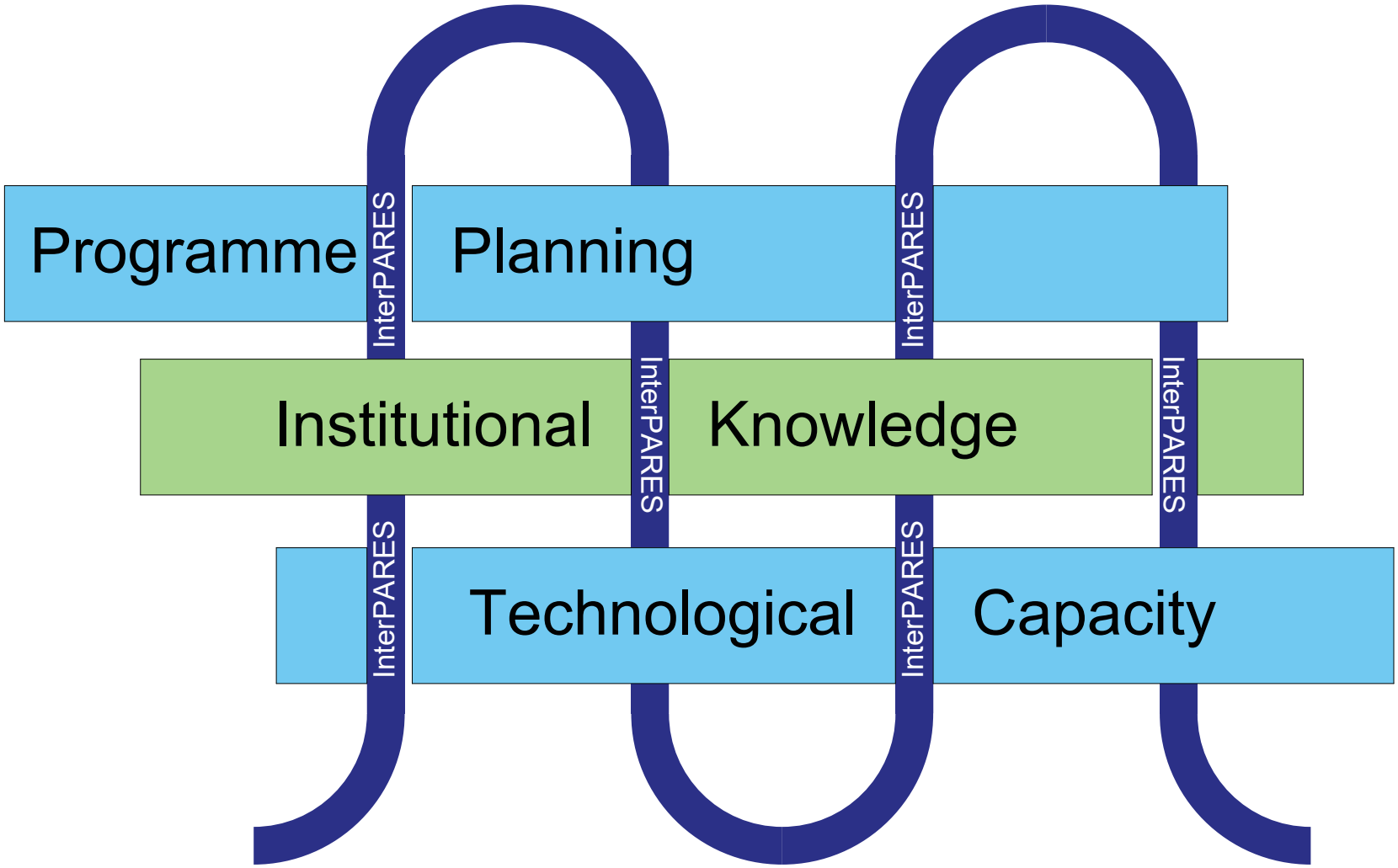
InterPARES Project
Glenn Dingwall, City of Vancouver Archives
Researcher – TEAM Canada

Presentation Structure

- Context and Background
  - City of Vancouver
  - Archives and Records Programmes
- Current Digital Records Challenges
- COV and InterPARES

Introduction

Weft: the common thread in a woven fabric that provides the structure, around which are interwoven the threads that form the pattern of the whole


Programme
Planning
Institutional
Knowledge
Technological
Capacity
InterPARES
InterPARES
InterPARES
InterPARES
InterPARES

Digital Preservation Threads

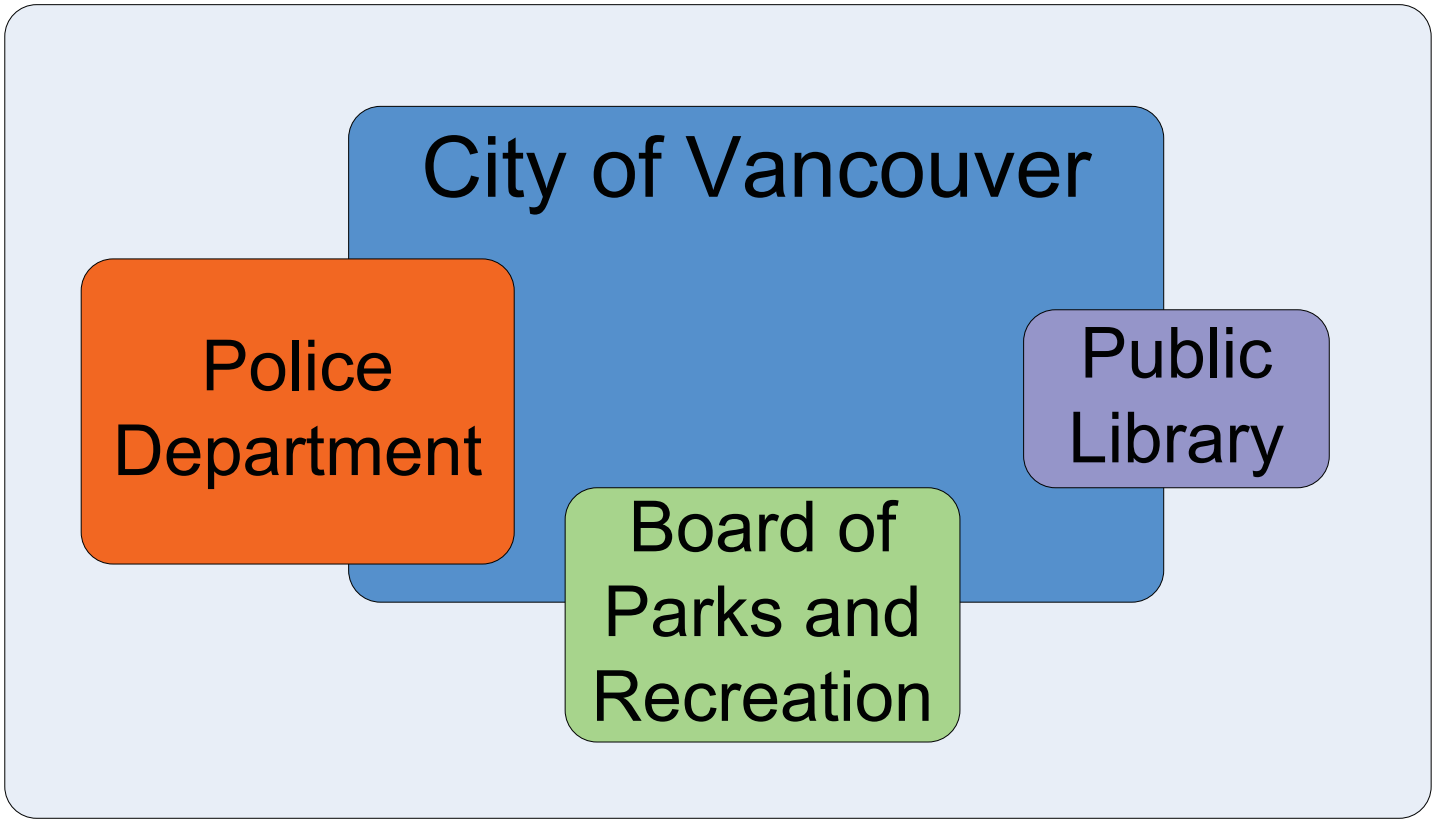
1. Programme Planning and Policy Infrastructure
2. Institutional Knowledge and Expertise
3. Technological Capacity

Some Context…

Vancouver Canada

- Incorporated 1886
- Population ~580k (2M in Metro Region)
  - Largest City in Region and Province
  - 3rd Largest City in Canada

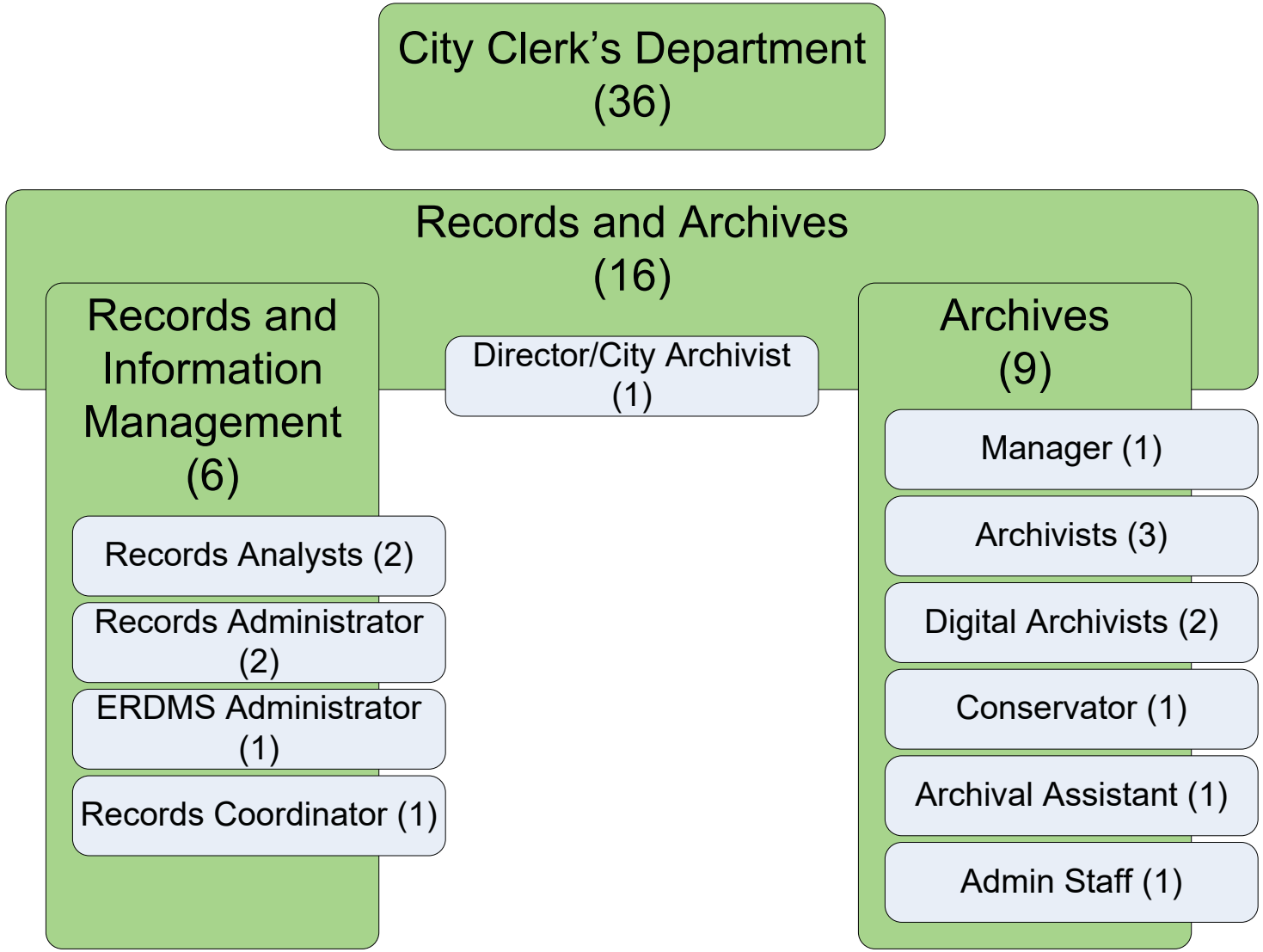
Vancouver – Juridical Components



COV Functional Responsibilities

- Community Centres
- Cultural Services
- Fire Protection
- Land Use Planning & Development
- Libraries
- Parks, Recreation & Public Spaces
- Permits, Licenses & Inspections
- Policing
- Social Planning
- Streets & Traffic
- Utilities
- Waste Management & Environmental Protection

Records and Archives – Org Context



Digital Preservation Challenges

- Total Archives Mandate
- Public Records Obligations
  - to the City itself
    - continuing administrative and operational needs
    - corporate memory
  - to the Citizens of Vancouver
    - to establish their rights against the government
    - to hold government accountable for its actions
- Private Records
  - Preserving the historical trace of life in Vancouver

Public Records

- Public Records
  - ERDMS
  - Unstructured network ‘shared drives”
  - Enterprise Resource Planning Systems (SAP)
  - Geographic Information Systems (VanMap)
  - Over 250 others on the IT System Inventory

Private Records

- Virtually no ability to control environment in which private records are created and kept
- Potential donations range from:
  - small donations containing well known formats
    - digital images
    - office documents
  - Large donations with diverse and (sometimes) obscure formats, such as...

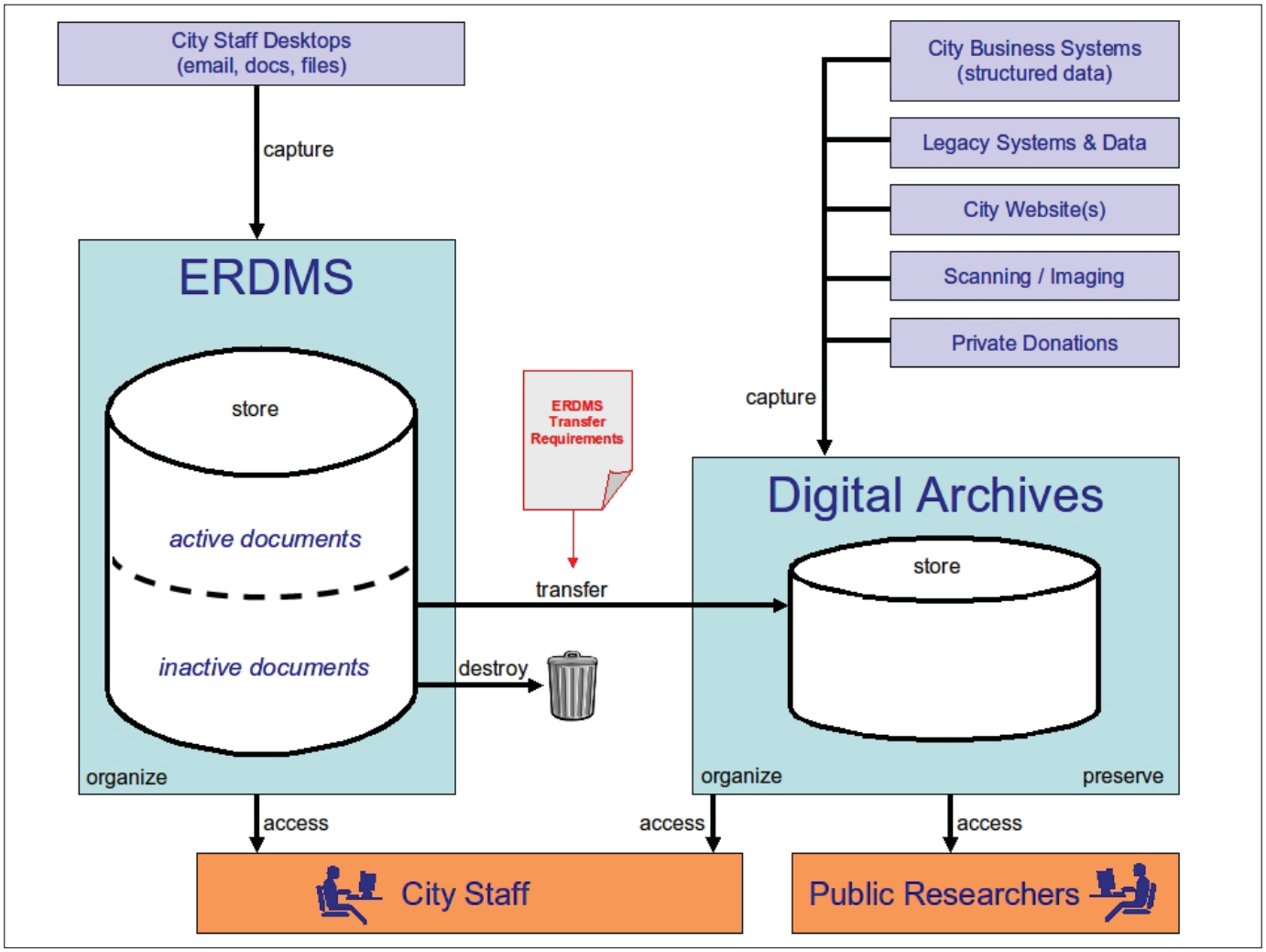
Records of the 2010 Games

vancouver 2010

vancouver 2010
PARALYMPIC GAMES
JEUX PARALYMPIQUES

InterPARES Project
Glenn Dingwall, City of Vancouver Archives
Researcher – TEAM Canada

Record Producers & Consumers



The Threads

1. Programme Planning and Policy Infrastructure
2. Institutional Knowledge and Expertise
3. Technological Capacity

Programme Planning

Effective digital preservation:

- begins early in the records life-cycle
- distributes responsibilities throughout the organization – various roles and at all levels
    - Requires that management create and support policies
    - Requires that records actors comply with policies, and have the ability to comply with policies

Utility of COP Model

- Linking activities throughout the life-cycle
- Linking high-level activities to low-level outputs
- Communicating requirements throughout the organization:
  - Vertically
  - Horizontally

Examples in Practice

- Linking ERDMS funding to Digital Archives funding
- Embedding preservation requirements in ERDMS RFP
- Linking Business Process Analysis to Digital Preservation through Preservation Plans attached to Classification/Disposition schedules
- Linking record analysis to IT systems procurement

Institutional Capacity

- Beneficial relationship of COV to UBC SLAIS & InterPARES Project
- Participation in case studies creates awareness in the City
- Participation in project also allows us to see what similar institutions are dealing with, gauge our relative level

InterPARES2/Case Study 24 – Preserving the VanMap GIS

Diverse sources of information, source formats, frequency of updates, significance of different elements to authentically conveying the intended message

CS24 provided roadmap to how to go about preserving VanMap, but digital preservation programme not mature enough at the time to act on findings and implement recommendations

IP2/CS24 Outcomes

1) Preserve content from a GIS environment.
   - Extract data for each GIS component
   - Data layers come from different storage resources
   - Demonstrate can assemble components into a viable GIS system
2) Preserve information about changes to the GIS environment.
   - Test whether a database can track changes to the GIS environment
   - From database, re-create GIS configuration file for requested time
   - Demonstrate application in a GIS system
3) Preserve snapshots of the VanMap system.
   - Demonstrate migration of GIS system into alternate technology

Technical Capacity – Digital Archives Project

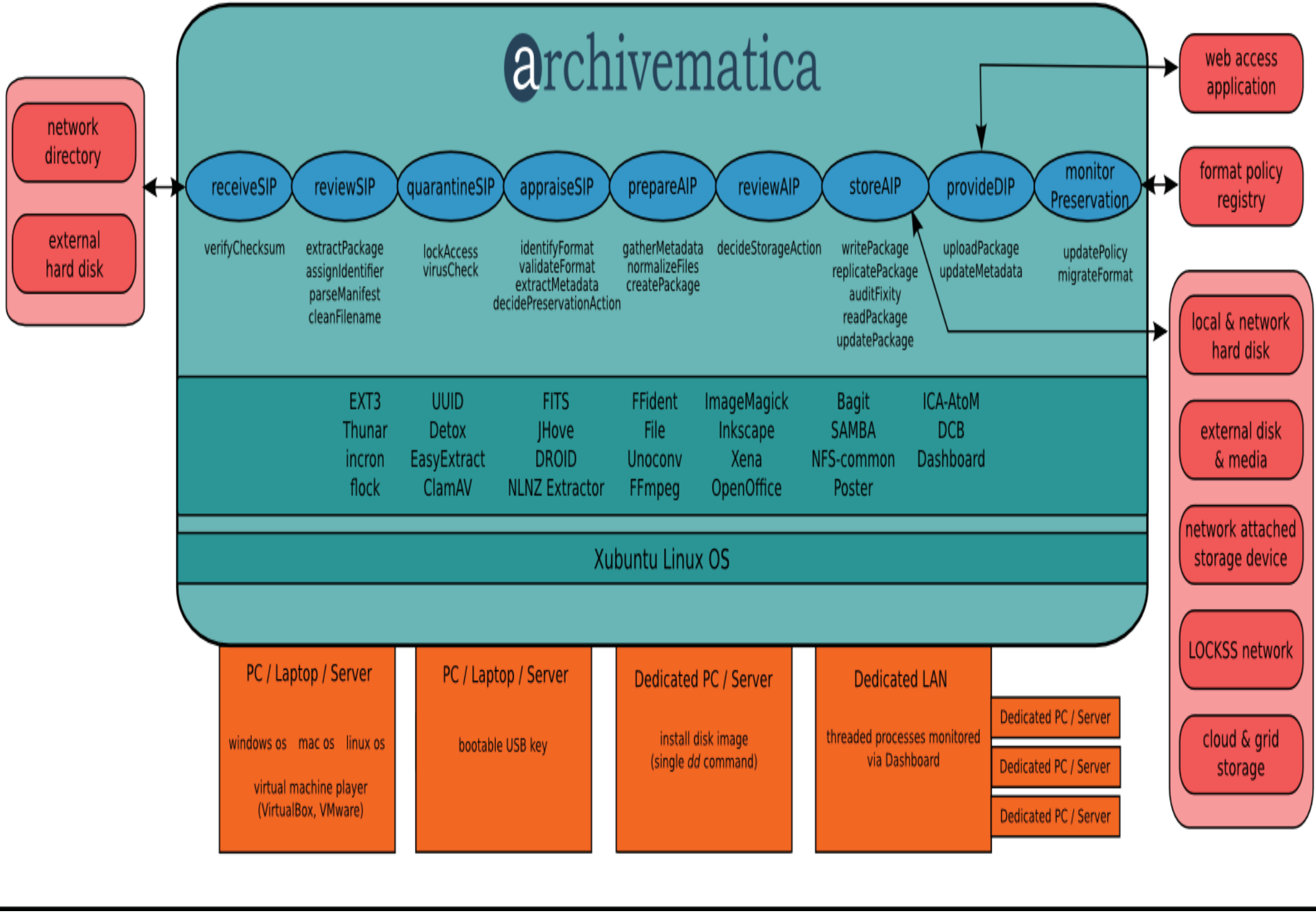
Goal: Create an digital preservation environment capable of acquiring, preserving and providing access to digital records.

Requirements Sources Include:

- OAIS Reference Model
- InterPARES findings
- UNESCO –*Towards an Open Source Repository*
- METS
- PREMIS

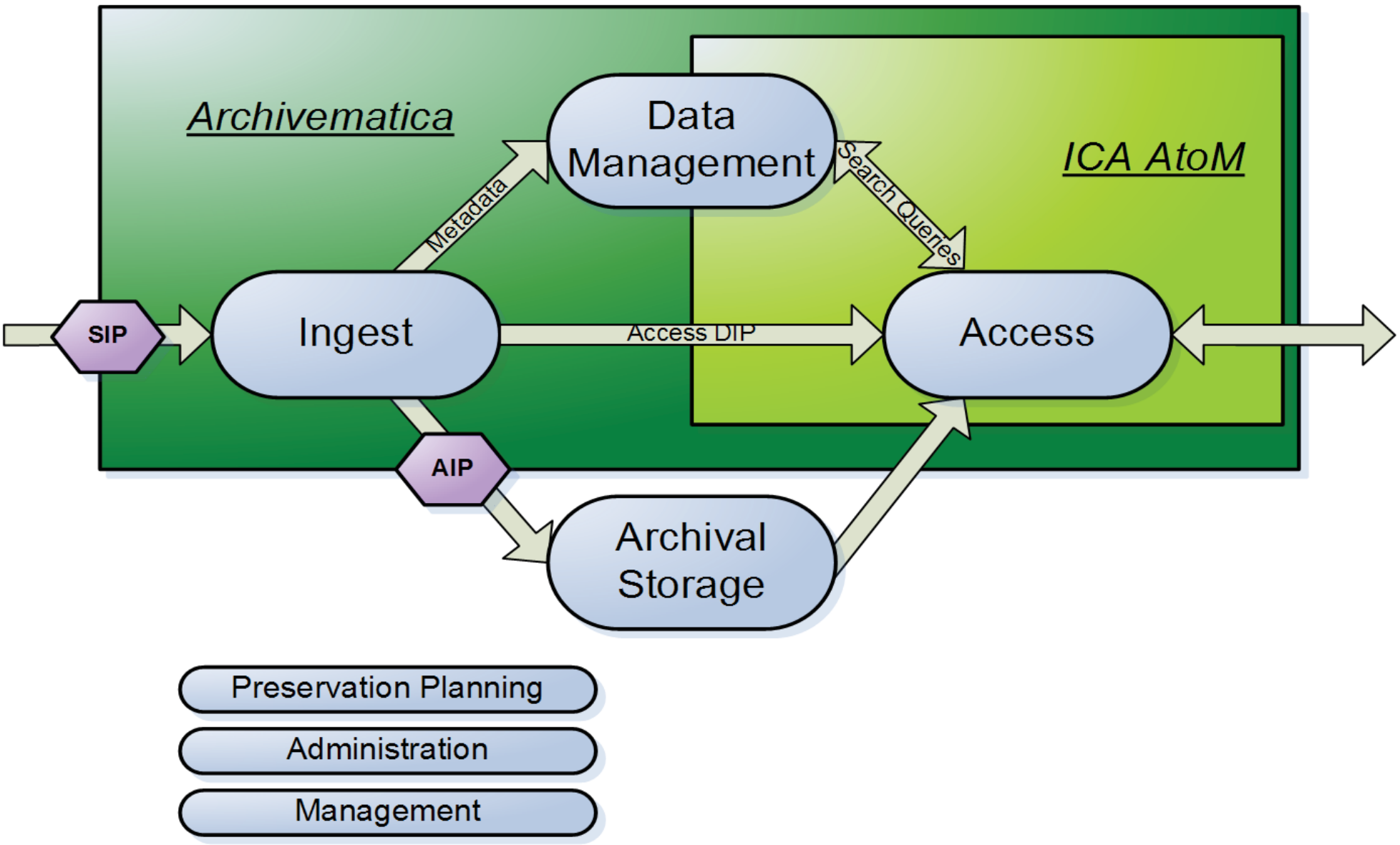
Digital Archives Project Timeline

	Phase 1	Phase 2
Timeline:	Nov 2008-Oct 2009	Nov 2009-Dec 2010
Focus:	Public records (ERDMS)	Private records (VANOC)
Funding:	~$150,000	~$580,000
Source:	ERDMS Project Funds	Olympic Legacy Fund

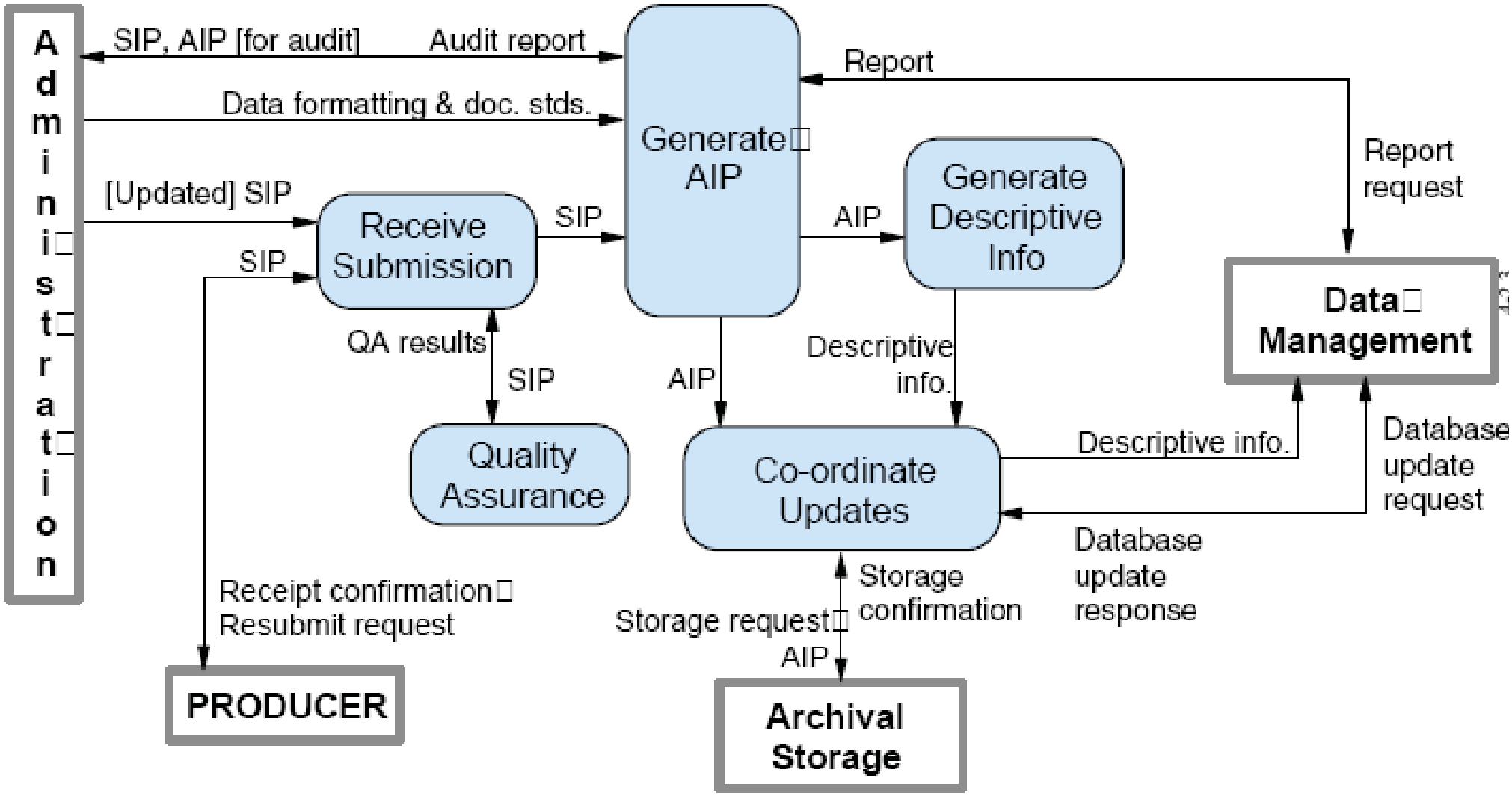

archivematica
network directory
external hard disk
receiveSIP
reviewSIP
quarantineSIP
appraiseSIP
prepareAIP
reviewAIP
storeAIP
provideDIP
monitor Preservation
verifyChecksum
extractPackage
assignIdentifier
parseManifest
cleanFilename
lockAccess
virusCheck
identifyFormat
validateFormat
extractMetadata
decidePreservationAction
gatherMetadata
normalizeFiles
createPackage
decideStorageAction
writePackage
replicatePackage
auditFixity
readPackage
updatePackage
uploadPackage
updateMetadata
updatePolicy
migrateFormat
web access application
format policy registry
local & network hard disk
external disk & media
network attached storage device
LOCKSS network
cloud & grid storage
EXT3
Thunar
incron
flock
UUID
Detox
EasyExtract
ClamAV
FITS
JHove
DROID
NLNZ Extractor
FFident
File
Unoconv
FFmpeg
ImageMagick
Inkscape
Xena
OpenOffice
Bagit
SAMBA
NFS-common
Poster
ICA-AtoM
DCB
Dashboard
Xubuntu Linux OS
PC / Laptop / Server
windows os mac os linux os
virtual machine player (VirtualBox, VMware)
PC / Laptop / Server
bootable USB key
Dedicated PC / Server
install disk image (single dd command)
Dedicated LAN
threaded processes monitored via Dashboard
Dedicated PC / Server
Dedicated PC / Server
Dedicated PC / Server

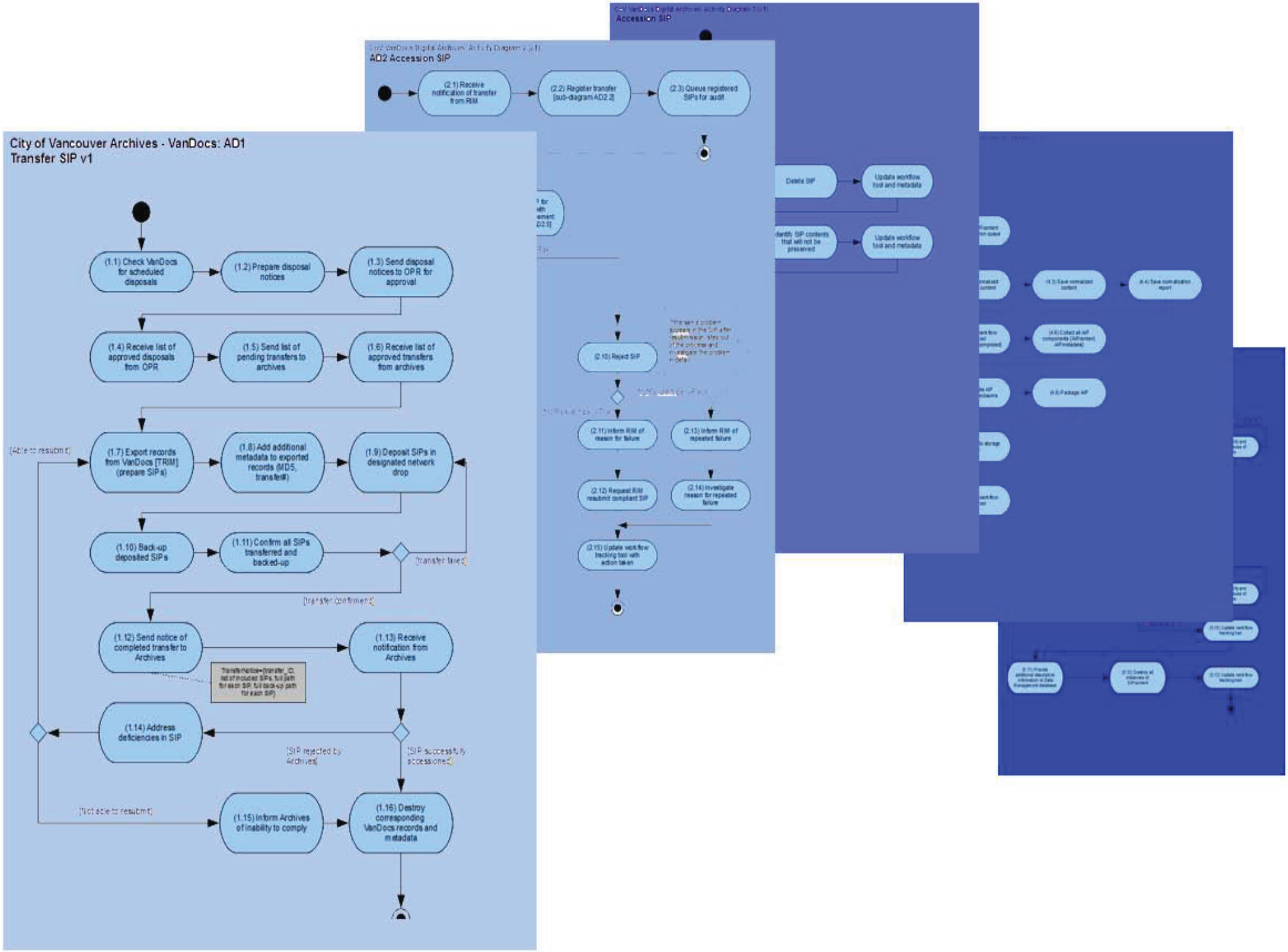
IP3/CS16 Requirements Gap Analysis

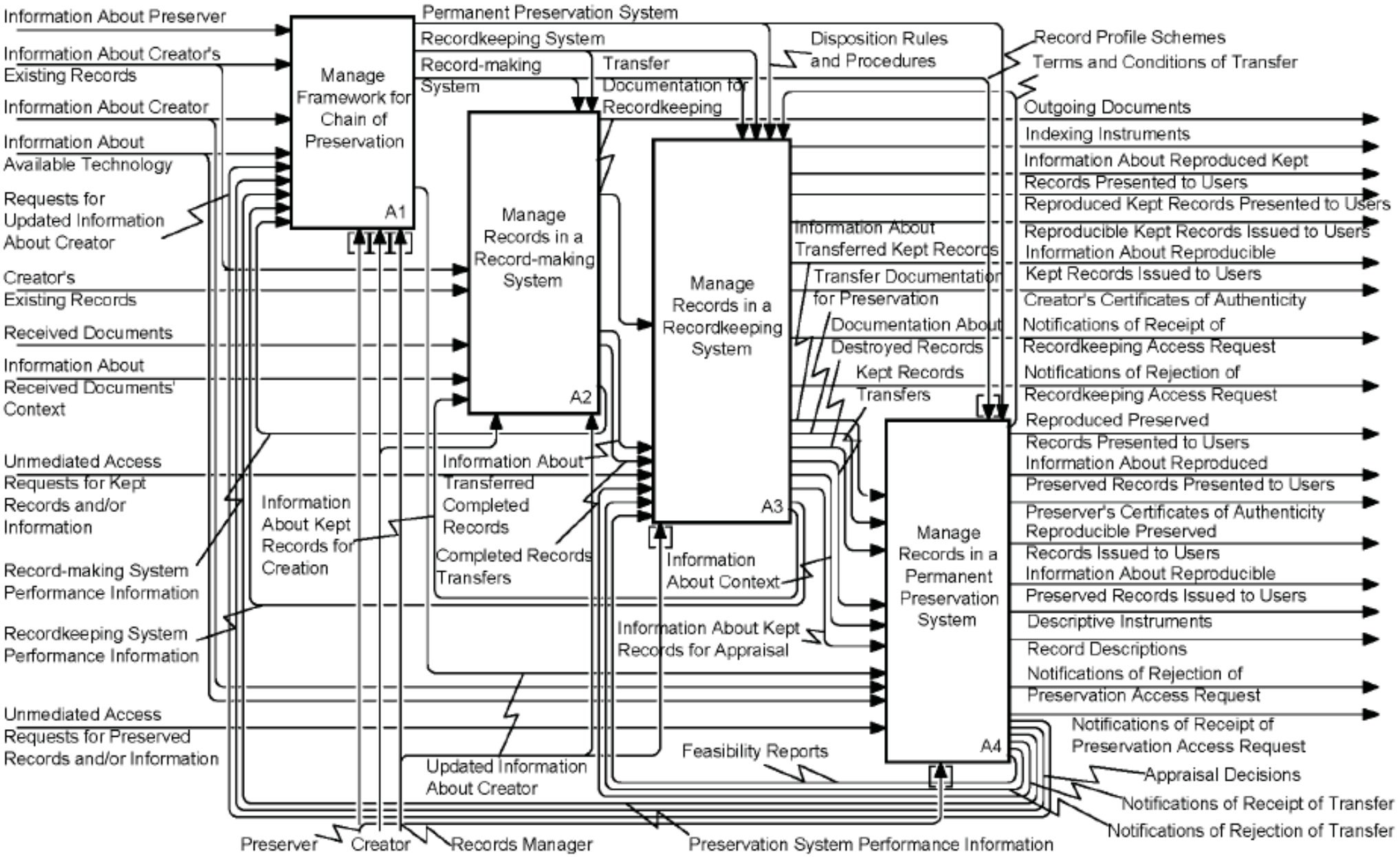
- Spring 2009
- Reviewed OAIS derived system requirements (UML diagrams) against IP2 Chain of Preservation Model (COP)
- OAIS Components
  - Ingest and Adjacent
- COP Components
  - A4 Manage Records in a Permanent Preservation System


Archivematica
Data Management
ICA AtoM
Metadata
Search Queries
SIP
Ingest
Access DIP
Access
AIP
Archival Storage
Preservation Planning
Administration
Management

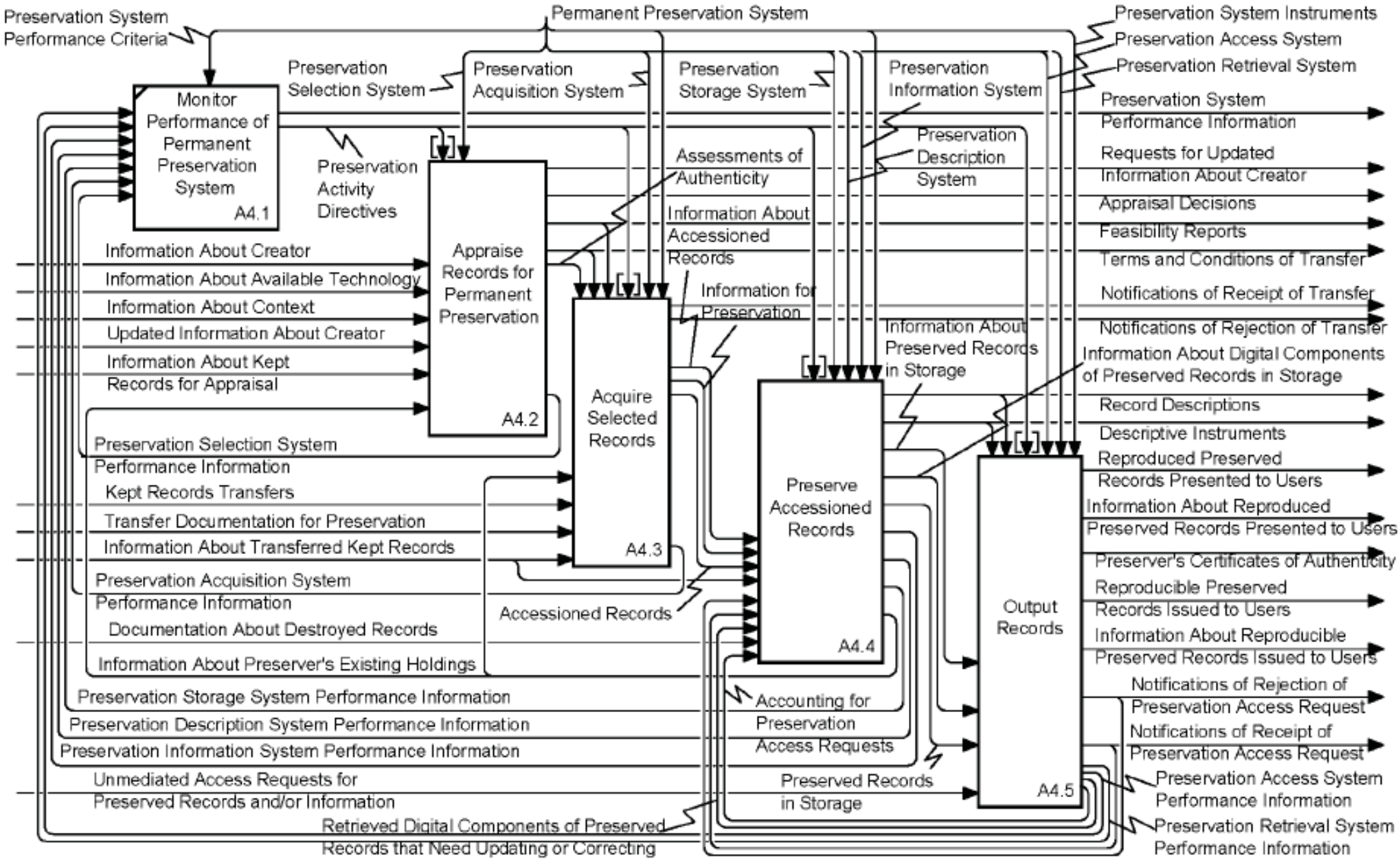
Ingest


Administration
SIP, AIP [for audit]
Audit report
Data formatting & doc. stds.
Generate AIP
Report
Report request
[Updated] SIP
Receive Submission
SIP
SIP
Generate Descriptive Info
AIP
Data Management
QA results
SIP
AIP
Descriptive info.
Quality Assurance
Co-ordinate Updates
Descriptive info.
Database update request
Database update response
Storage confirmation
Storage request
AIP
Receipt confirmation
Resubmit request
PRODUCER
Archival Storage


City of Vancouver Archives - VanDocs: AD1
Transfer SIP v1
(1.1) Check VanDocs for scheduled disposals
(1.2) Prepare disposal notices
(1.3) Send disposal notices to OPR for approval
(1.4) Receive list of approved disposals from OPR
(1.5) Send list of pending transfers to archives
(1.6) Receive list of approved transfers from archives
[Able to resubmit]
(1.7) Export records from VanDocs [TRIM] (prepare SIPs)
(1.8) Add additional metadata to exported records (MD5, transfer#)
(1.9) Deposit SIPs in designated network drop
(1.10) Back-up deposited SIPs
(1.11) Confirm all SIPs transferred and backed-up
[transfer failed]
[transfer confirmed]
(1.12) Send notice of completed transfer to Archives
(1.13) Receive notification from Archives
(1.14) Address deficiencies in SIP
[SIP rejected by Archives]
[SIP successfully accessioned]
[Not able to resubmit]
(1.15) Inform Archives of inability to comply
(1.16) Destroy corresponding VanDocs records and metadata
AD2 Accession SIP
(2.1) Receive notification of transfer from RIM
(2.2) Register transfer [sub-diagram AD2.2]
(2.3) Queue registered SIPs for audit
(2.10) Reject SIP
(2.11) Inform RIM of reason for failure
(2.12) Request RIM resubmit compliant SIP
(2.13) Inform RIM of repeated failure
(2.14) Investigate reason for repeated failure
(2.15) Update workflow tracking tool with action taken
Accession SIP
Delete SIP
Update workflow tool and metadata
Update workflow tool and metadata



NODE:	TITLE:	NUMBER:
A0	Manage Chain of Preservation	



NODE:	TITLE:	NUMBER:
A4	Manage Records in a Permanent Preservation System	

Gap Analysis Findings

2009 Report identified apparent deficiencies related to:

- Appraisal
- Arrangement
- Authenticity

Recommended COP as tool for conducting a metadata analysis

Metadata Analysis

- PREMIS-COP mapping
- AIP Profile
- Top Down Analysis by IP Researchers and GRAs
- Bottom-up Analysis by COV/Artefactual
- Results to be incorporated into Archivematica v0.7 (scheduled Release 4th quarter of 2010)

Links

City of Vancouver Archives

vancouver.ca/archives

Digital Archives Project Wiki

artefactual.com/wiki/index.php?title=Vancouver_Digital_Archives

Archivematica Wiki

archivematica.org/wiki