

Distributed
Archiving & Preservation System
Système de Préservation
et d'Archivage Réparti (S.P.A.R.)

Agenda

- General context
 - The BnF
 - Producers and users => channels
- Methodology
 - OAIS model
 - Analysis
- Solution applied
 - SPAR-Infrastructure
 - SPAR-Realization
- Status

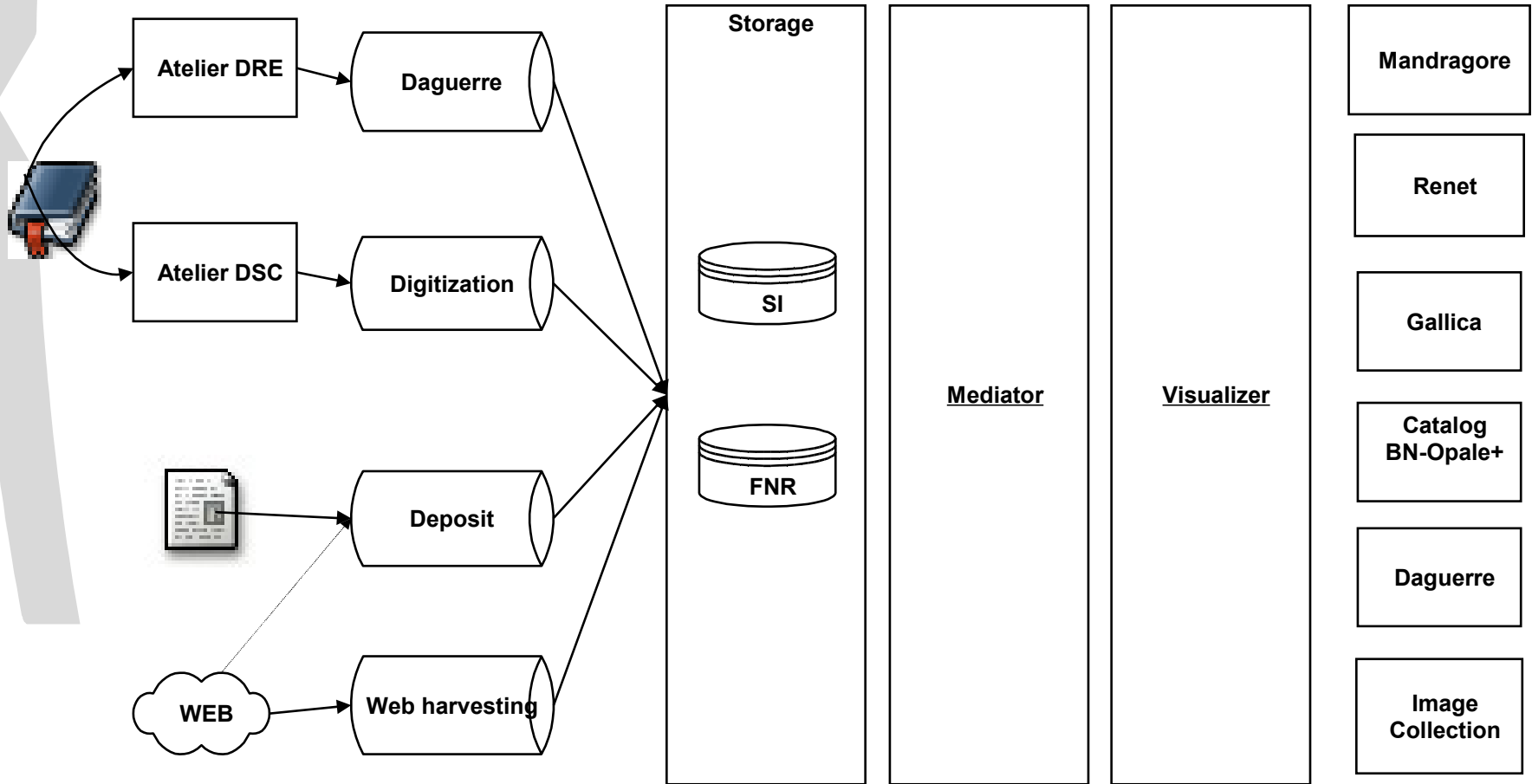
■ Missions :

- to build up the collections,
- to preserve and communicate them to the public,
- to produce a reference catalog,
- to cooperate with other institutions,
- to participate to research programs.

Legal deposit :

- legal deposit since 1537 for printed materials
- 1648: engravings and maps
- 1793: musical scores
- 1925: photos
- 1938: phonograms
- 1941: posters
- 1975: videograms and multimedia documents
- 1992: audiovisual and electronic documents
- 2006: Web legal deposit

Digital domain: actual status



Definition of a channel

- Build on the relation between the digital objects and the archival system, independently of any given organization.
- Defined by the mutual requirements:
 - From the point of view of the archival system with respect to the producer
 - ❖ Preliminary negotiation (possible or not)
 - ❖ Legislative and lawful aspect (required or not)
 - ❖ Access conditions
 - From the point of view of the producer with respect to the archival system
 - ❖ Data fixity (modification, version, deletion or not)
 - ❖ Permanency of the data (data, metadata, structures and representations)
 - ❖ Accessibility (direct / through an application, immediate or differed)

- Conservation digitization
- Reproduction digitization
- Automatic legal deposit (surface Web)
- Negotiated legal deposit (dark Web, regional press)
- Administrative production
- Deposit / Third party archiving
- Acquisition / Donation

Agenda

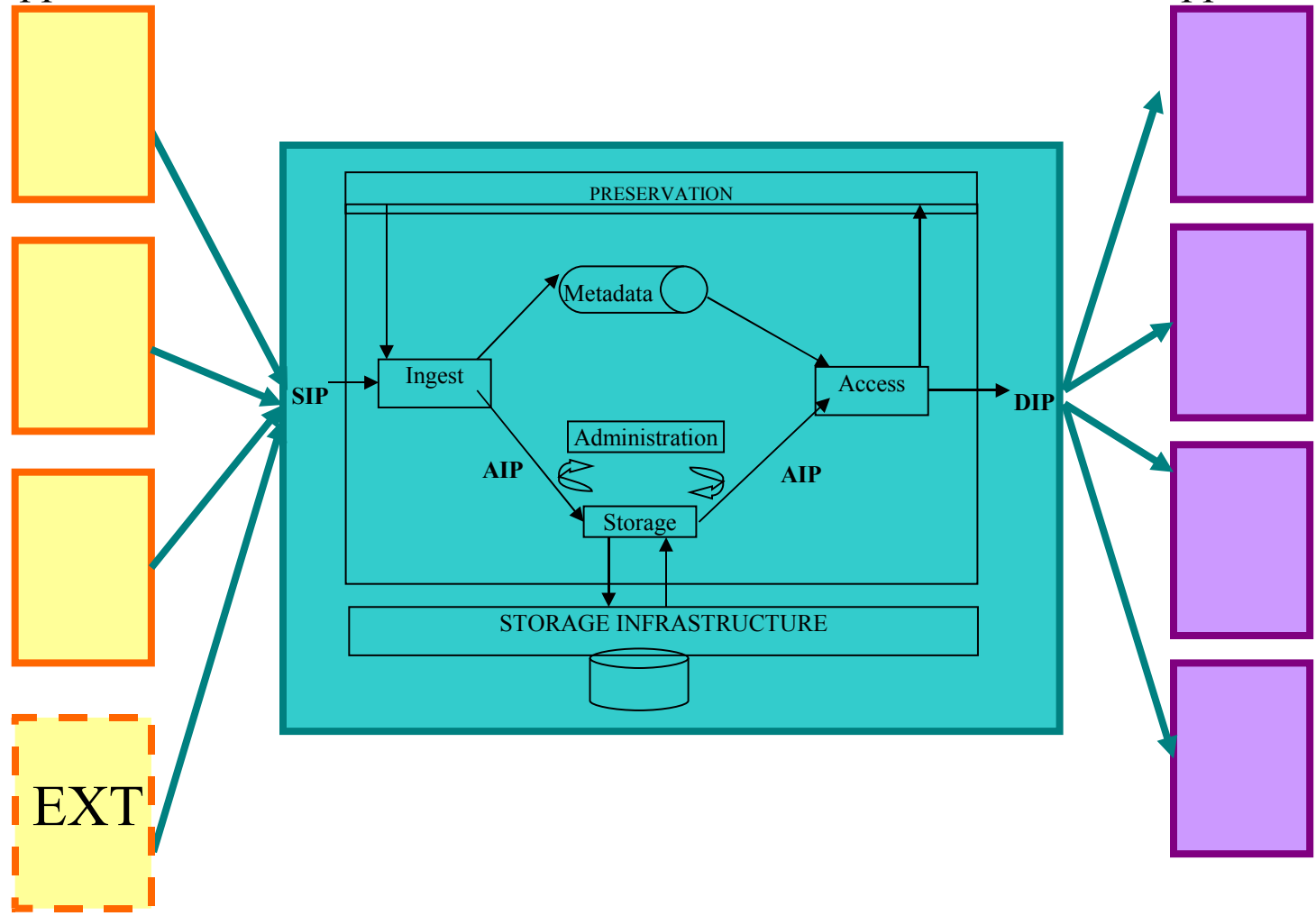
- General context
 - The BnF
 - Producers and users => channels
- Methodology
 - OAIS model
 - Analysis
- Solution applied
 - SPAR-Infrastructure
 - SPAR-Realization
- Status

- Open Archival Information System (OAIS)
 - *Reference Model for an Open Archival Information System (OAIS)*. Blue Book. Version : 1st January 2002.
 - International standard : **ISO 14721:2003**. Space data and information transfer systems - Open archival information system - Reference model
 - ❖ Translation in French by CNES and BnF

Digital domain: the future

Production applications

Diffusion applications



- Needs definition (library wide groups)
 - G1: functions
 - G2: user communities => channels
 - G3: information models (metadata and packaging information)
 - G4: risk management
 - G2D: digital right management
- Technical study of possible solutions

- Technical study of possible solutions through 2 steps evaluation
 - Coarse grain evaluation : large number of solutions (20) / limited number of criteria
 - Fine grain evaluation :
 - ❖ 6 solutions
 - ❖ large number of BnF specific criteria (> 100)

=> Used for elaborating the RFP

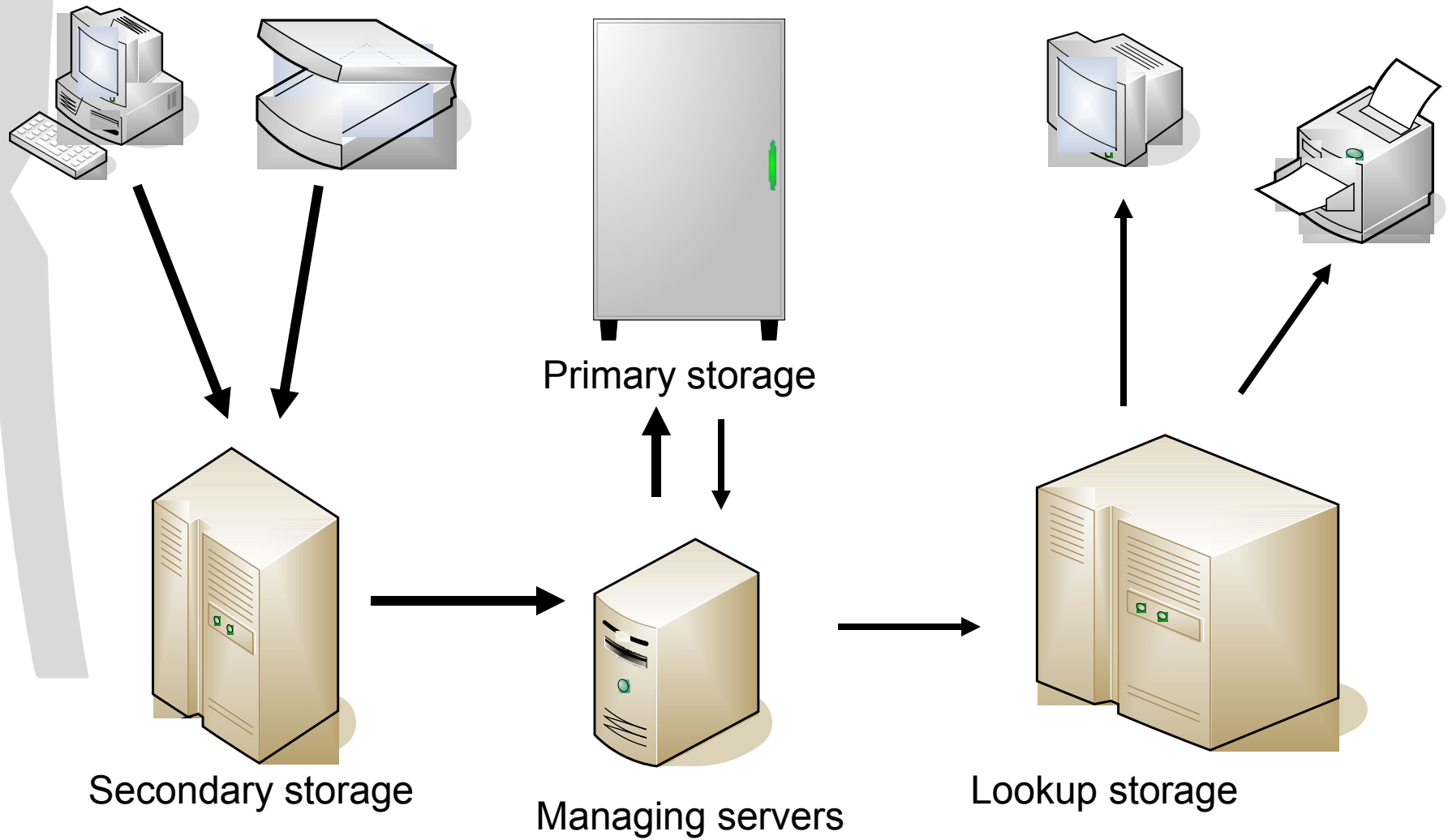
SPAR: the main requirements

- Technology: compatible with our infrastructure
- Permanency: commitment on durable technologies => OpenSource
- Customizability: extensible in a declarative way
- Maintainability: the evolution of a component must be cost effective and seamless
- Modularity: at the level of the OAIIS entities, each component has a clear interface and is testable by itself
- Scalability and distributivity : at the level of the module
- Transparency to migration: migration of objects have no effect on access
- Hardware independency: “storage abstract service”

Agenda

- General context
 - The BnF
 - Producers and users => channels
- Methodology
 - OAIS model
 - Analysis
- Solution applied
 - SPAR-Infrastructure
 - SPAR-Realization
- Status

- **Primary storage**
 - Data archiving
- **Secondary storage**
 - Storage of the data before the archiving
- **Lookup storage**
 - Access cache
- **Managing Servers**
 - Data processing
 - Data store / restore
 - Media supervising





- Tape libraries providing the function of storing the data

- **Openness**: compatibility with at least 3 different tape technologies
- **Availability** : retrieval of 50Mb of data in less than 40 seconds
- **Upgradeability** : volume, modularity

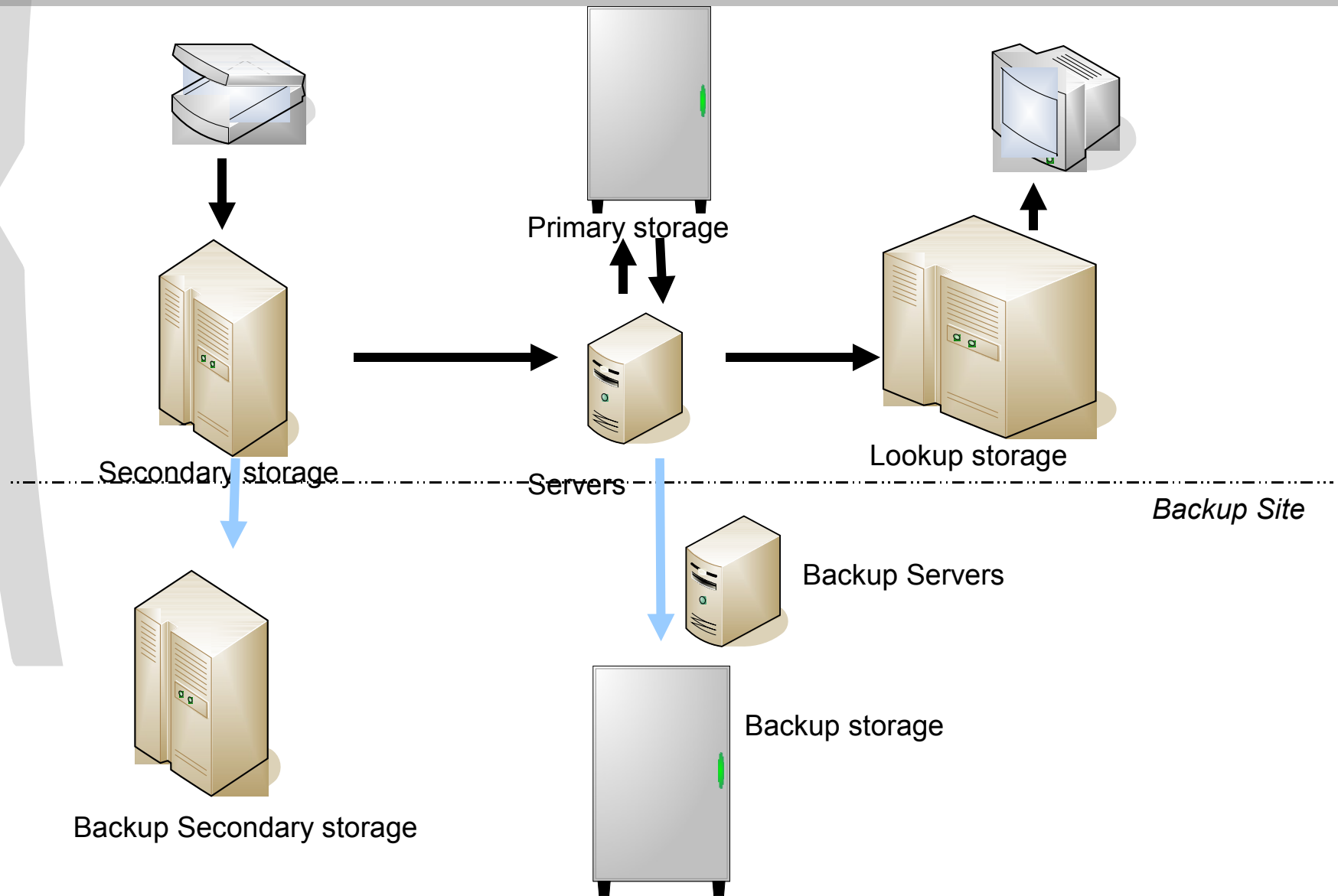
=> Sun StorageTek SL8500



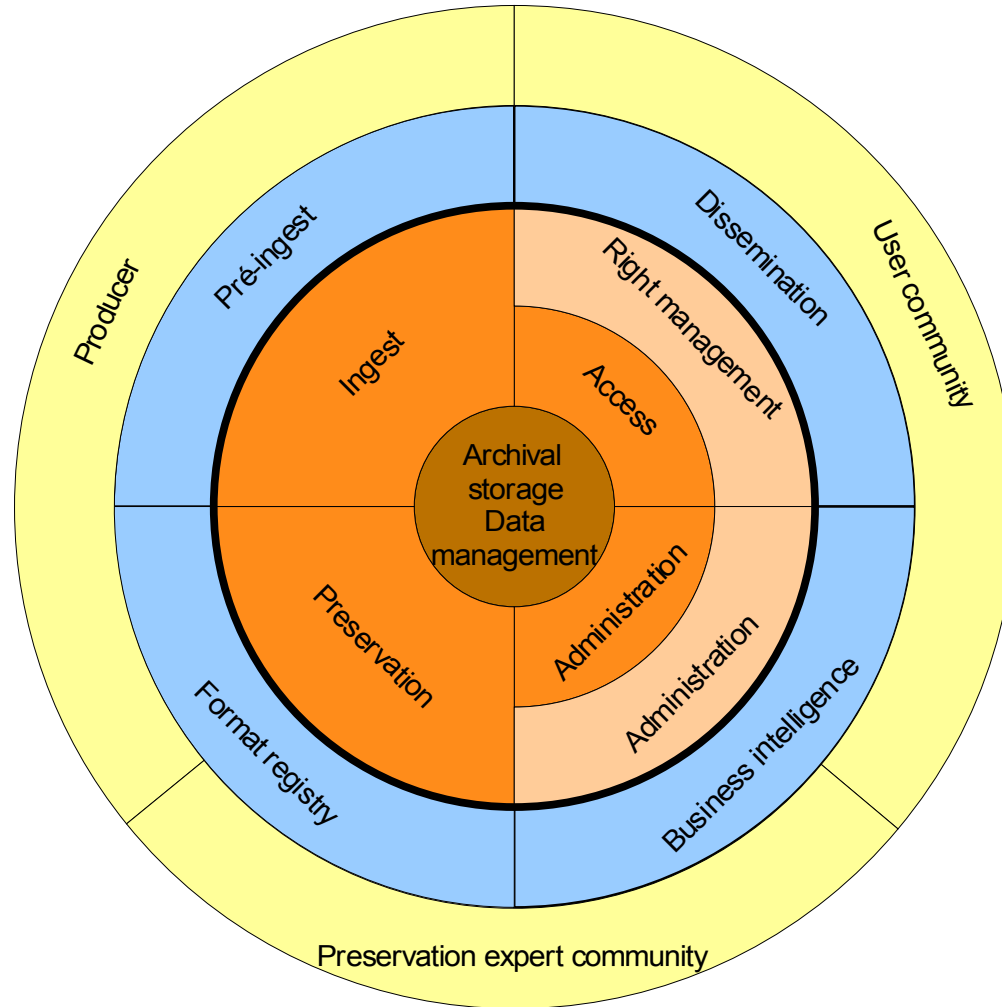
- Disk arrays for temporary data to will be archived or to be access
- Data provided by external providers to the system
- Speed of access

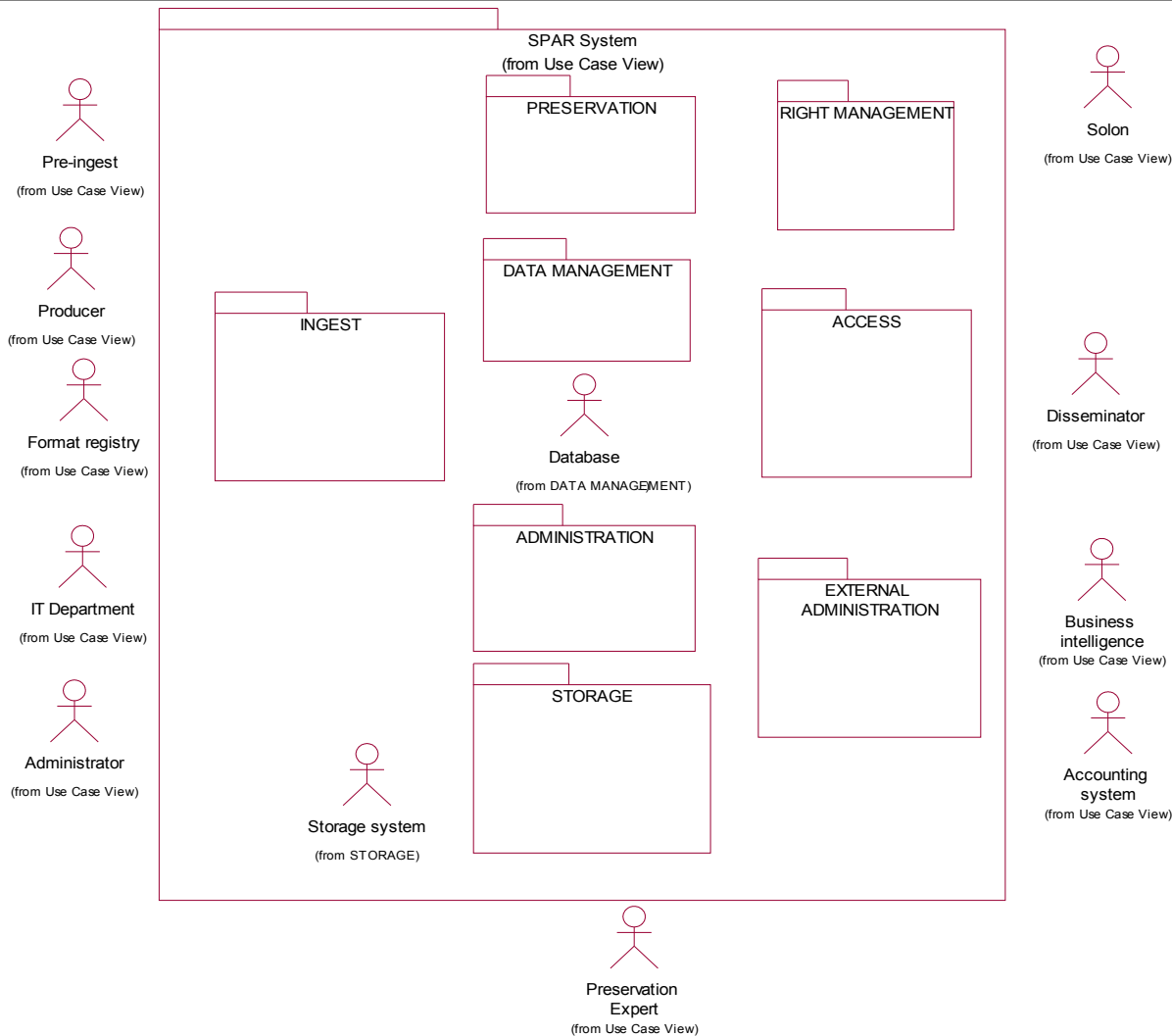
- **Openness:** CIFS (Windows) and NFS (Unix) protocols
- **Availability :** redundancy (electrical, network, RAID, ...)
- **Upgradeability :** volume, modularity

=> **NAS Sun Stk FLX 380**



SPAR-Realization : conceptual model





- External use case modeling
- Internal use case modeling

- Packaging through METS :
 - descriptive metadata: DC
 - technical metadata: depending on the channel (MIX, textMD, ...)
 - preservation metadata: PREMIS
 - right metadata: ODRL (for licenses)
- 4 levels of granularity (channel-independent):
 - set: collection (ex: whole collection of serials)
 - group: semantic unit (ex: a monograph, an issue of a serial)
 - object : part of the unit (ex: a page, an image)
 - file: bit-stream of a digital object

■ SPAR-Infrastructure

- initial contract in 2005
- last until 2009

■ SPAR-Realization

- RFP in progress
- 1st phase (14 months): to build the core and one channel
- 2nd phase (30 months): to iterate on the channels and improve the core

Questions ?