

# **Key changes of SPAR**

**(Scalable Preservation and Archiving Repository)**

**National Library of France  
Bibliothèque nationale de France**

September 17th, 2014

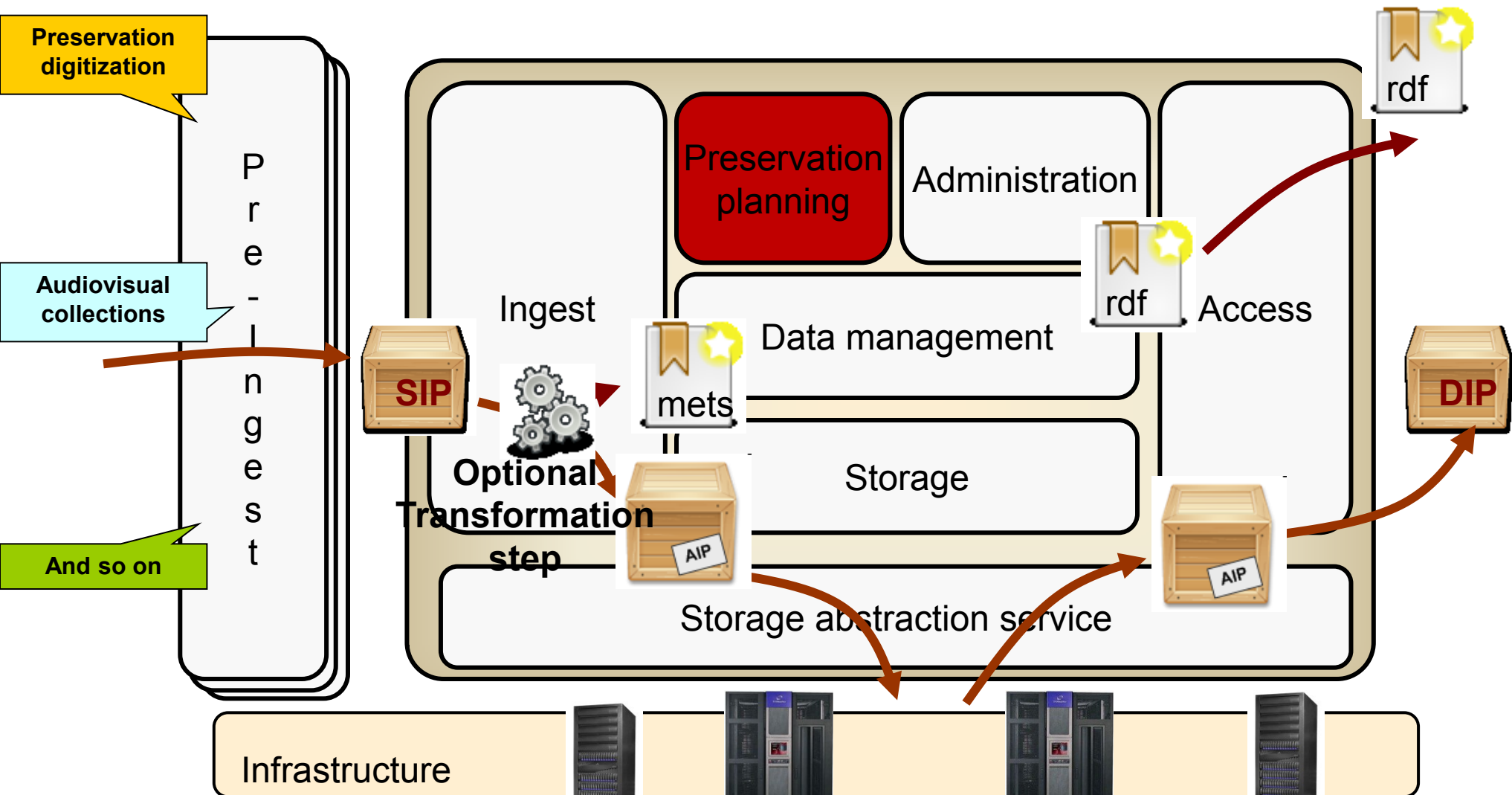
# Some facts about SPAR (at 09/08/2014)

- Operation begins: may 2010
  - Current workflow of BnF's mass digitization program
  - Distributed storage over 2 sites
- **3 030 450** packages archived
- 200 703 000 data-objects (files)
- **> 1.4PB** (=1000TB) of raw data
- **> 1.6 billion RDF triples** (elementary metadata)
- Very large manuscripts have been ingested (packages with size of 100GB)
- First technological migration made: new tape generation
- Available tracks: monographs, periodicals, still images, audio, video, web archiving, office documents, third-party archiving

# SPAR

{ BnF

## a generic repository solution at BnF



# Decomposition in tracks

- To deal with the variability and heterogeneity of the data, definition of **tracks**
  - build on the relation between the digital objects and the archival system, independently of any given organization:
    - Preservation digitization
    - Audiovisual legal deposit
    - Negotiated legal deposit (e-books, large posters,...)
    - Automatic legal deposit (surface Web)
    - Administrative production
    - Third party archiving
    - Acquisition / Donation
- + reference track**

# Channels currently in operation

- Reference channels (SLAs, formats, tools, tests)
- Preservation digitization: in doors and out doors
  - 1.6 million AIPs for 1.2 PB
- Audiovisual material
  - 62 000 AIPs for 57 TB
- Large posters legal deposit
  - 77 AIPs for 313 GB
- Web archiving
  - 1.3 million AIPs for 143 TB
- Third party archiving: 2 contracts
  - 56 000 AIPs for 6 TB

# Evolution in the organization

The library is building a new organization to interact with the Archive :

- census of preservation experts (expertise everywhere in the library)
- each track has a designated manager from the business units, to negotiate the SLAs and define the priorities
- the IT department is responsible for the data in “digital shelves”

Most importantly, preservation is now taken into account very early in the new projects