PASIG 2019

When digital preservation at BnF becomes business as usual...

National Library of France

Bibliothèque nationale de France

PASIG Mexico - 2019-02-14

BnF

Agenda

- Context of digital assets at BnF
- From archiving to preserving
- Crafting a new organization
- Sharing data and workflows
- Handling the volume
- Conclusion

The National Library of France

An autonomous public establishment

2 200 agents and dozens of professions, a budget of approx. 230 M€

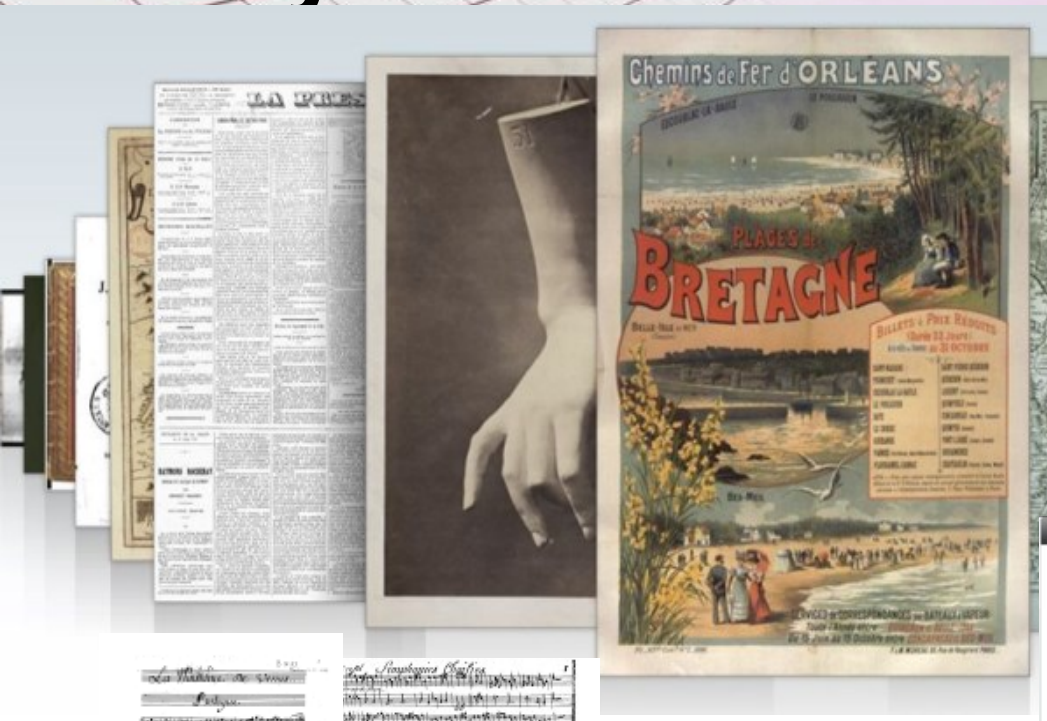
Local, national and international missions

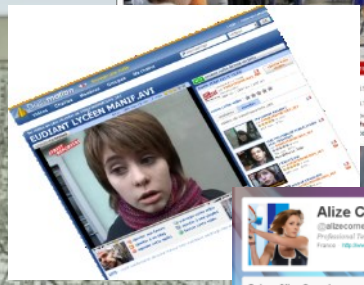
About 1M readers per year, 300 000 visitors to the exhibitions

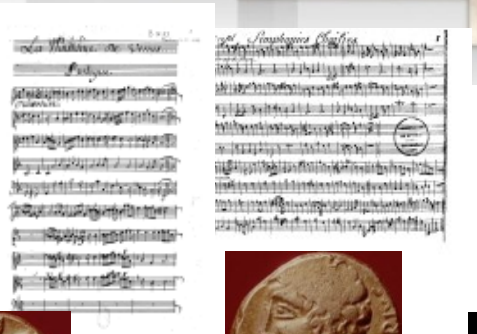
Large collections to manage

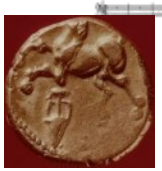
- More than 14M books
- More than 30M posters and photos
- More than 250.000 manuscripts
- More than 1M audiovisual material
- “French” Web legal deposit

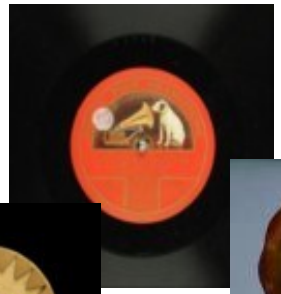
Variety of the collections



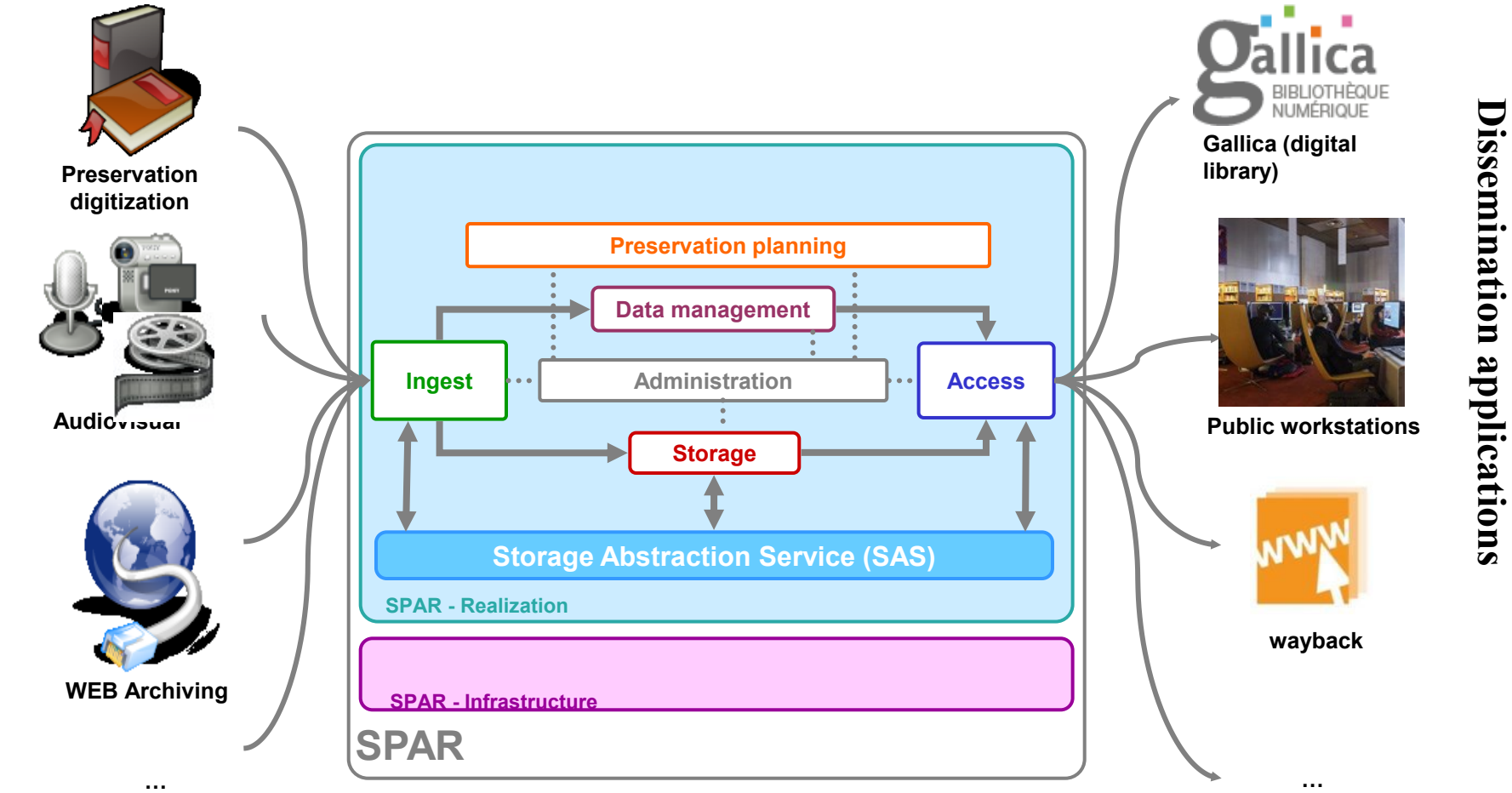








Digital preservation at BnF: SPAR



Some facts about SPAR (at 22/01/2019)

- Operation begins: may 2010
  - Current workflow of BnF's mass digitization program
  - Distributed storage over 2 sites
- 9 055 476 packages archived
- 386 830 827 data-objects (files)
- > 3.7PB (=3 785TB) of raw data
- > 5.2 billion RDF triples (elementary metadata)
- Very large manuscripts have been ingested (packages with size of 100GB)
- Two technological migration already made in 2012 and 2017: new tape generation
- Available tracks: monographs, periodicals, still images, audio, video, web archiving, office documents, ebooks, third-party archiving

Agenda

- Context of digital assets at BnF
- From archiving to preserving
- Crafting a new organization
- Sharing data and workflows
- Handling the volume
- Conclusion

Initially, a safe deposit

Entry chain

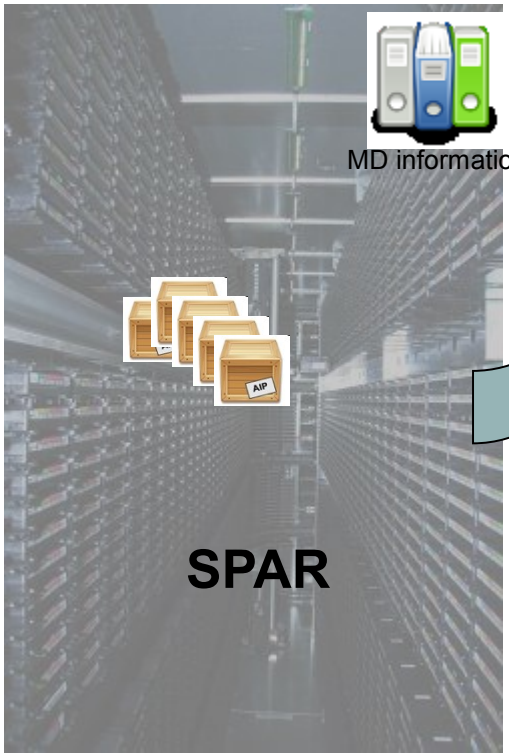
Storage

Ingest

Dissemination storage

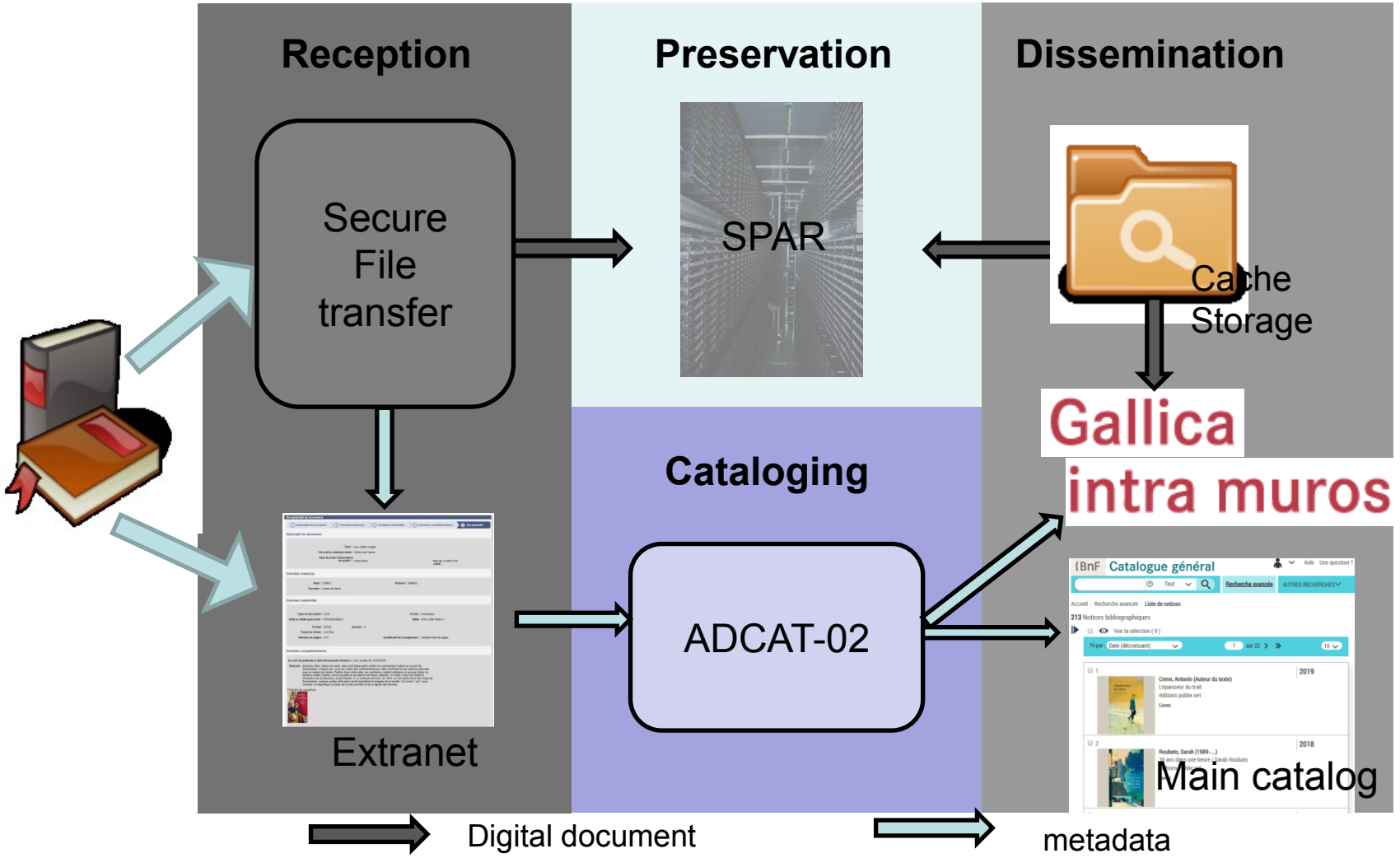
Dissemination

gallica
BIBLIOTHÈQUE NUMÉRIQUE



Initially, the information packages weren't disseminate directly

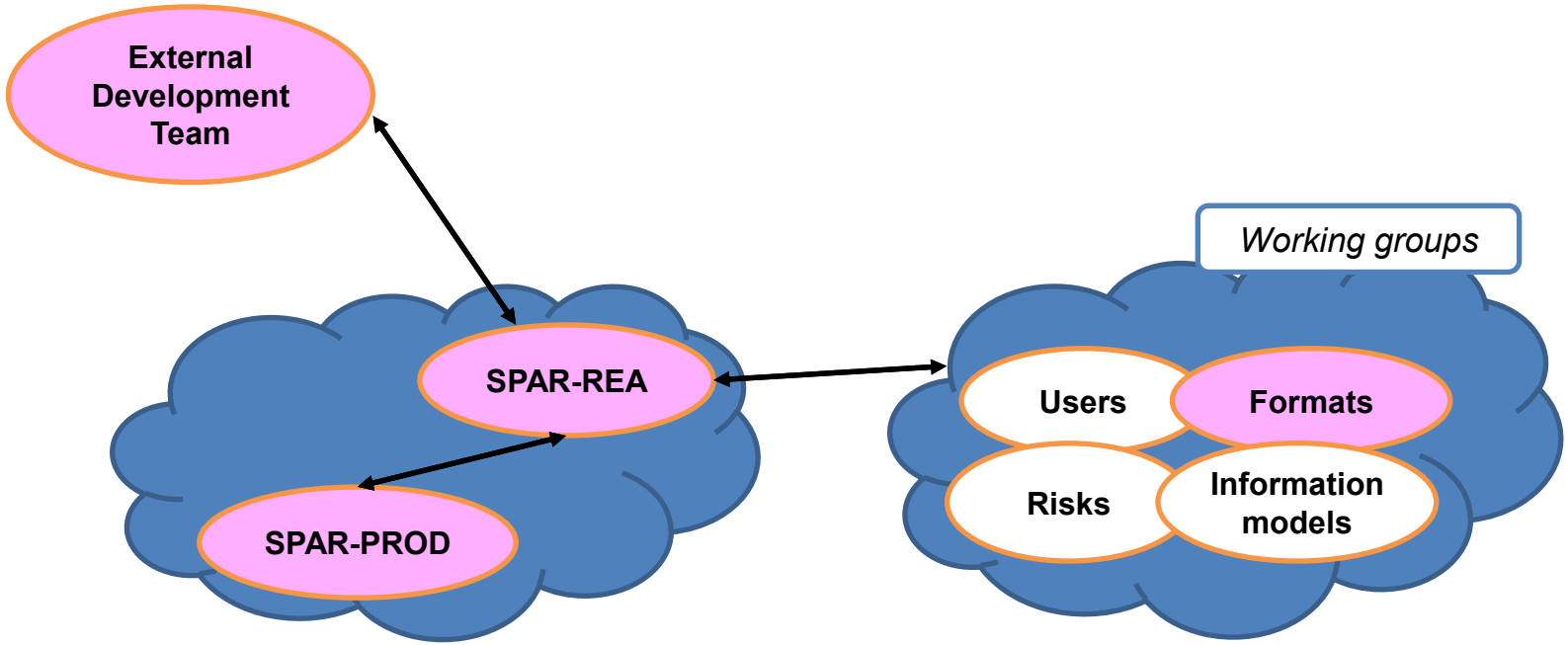
Now, digital stacks



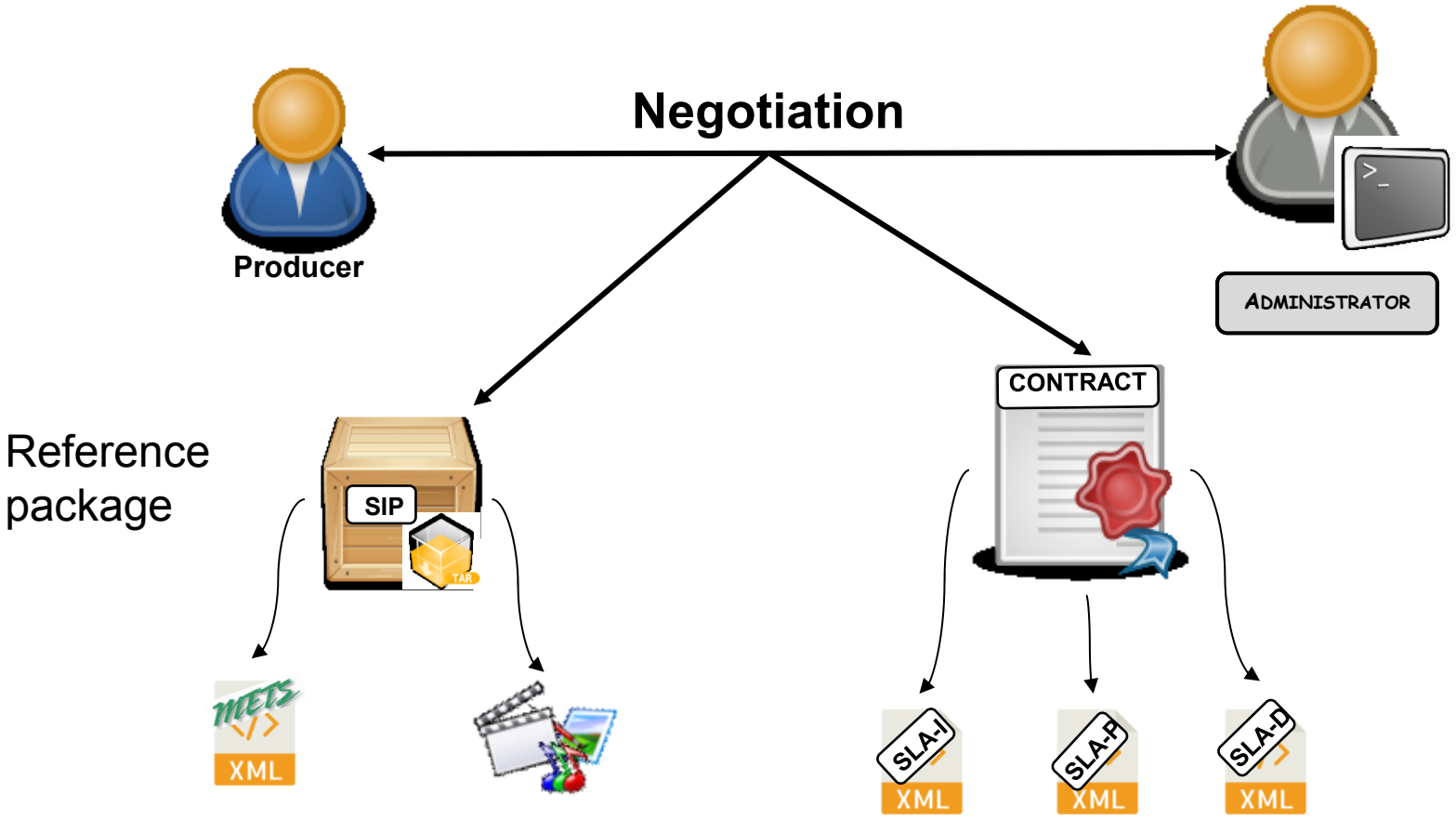
Agenda

- Context of digital assets at BnF
- From archiving to preserving
- Crafting a new organization
- Sharing data and workflows
- Handling the volume
- Conclusion

Initial project organization



Track negotiation: a contract



Reference package

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 the “digital stacks”

Most importantly, preservation is now part of every new digital project

People involved in preservation planning

TRACK MANAGER

ADMINISTRATOR

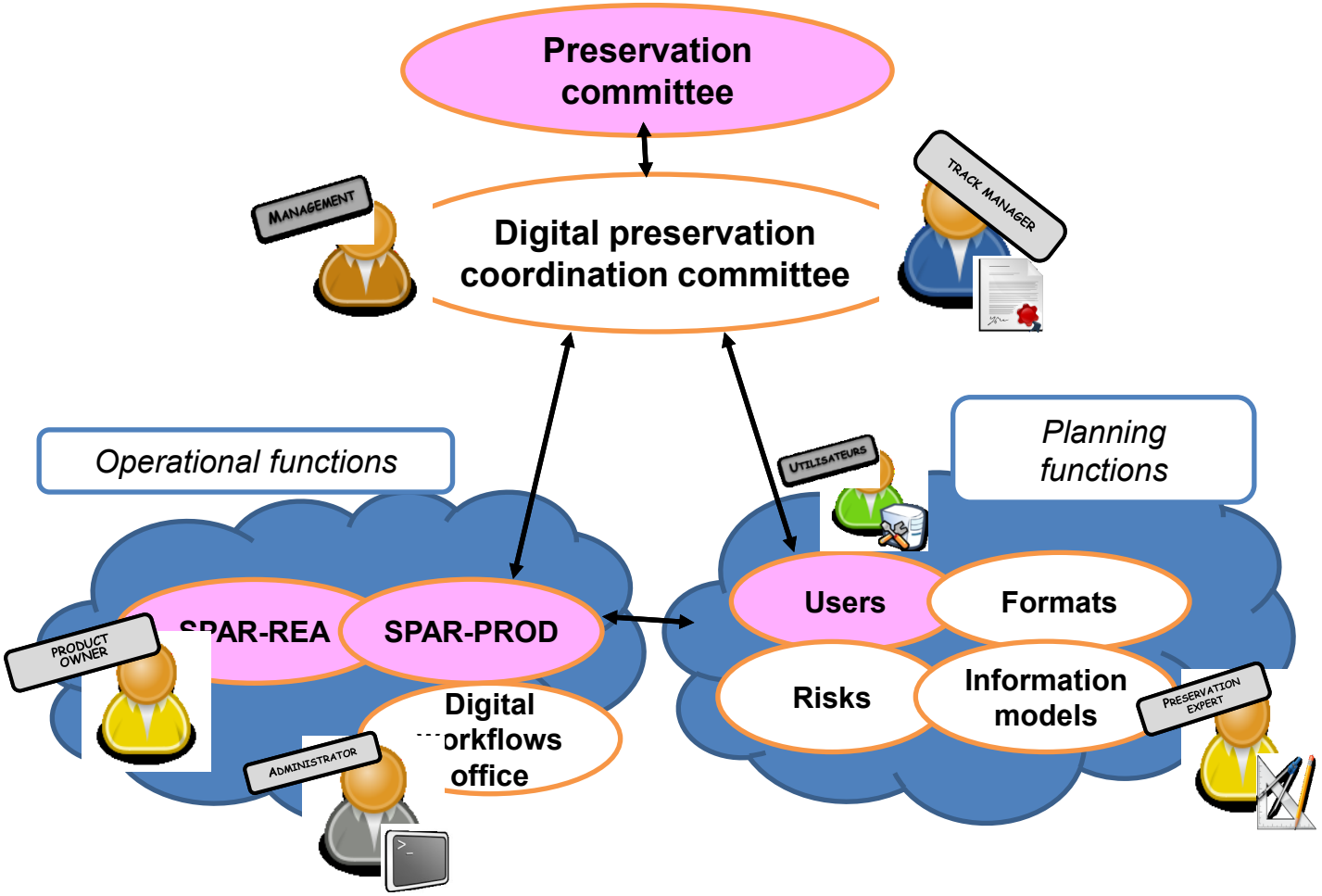
PRESERVATION EXPERT

DEVELOPER

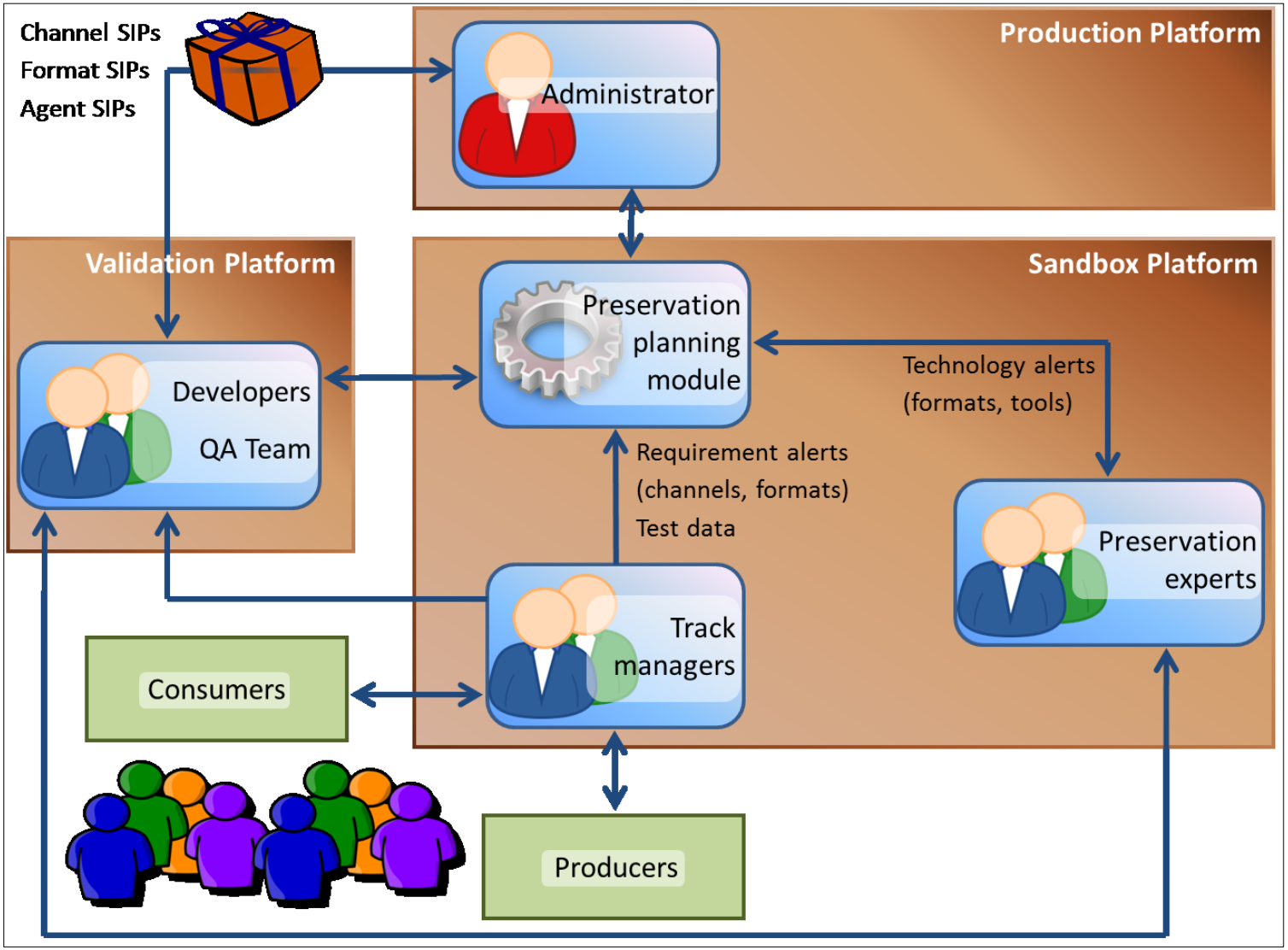
COLLECTION MANAGER

RISKS & EMERGENCY PLAN

Preservation organization



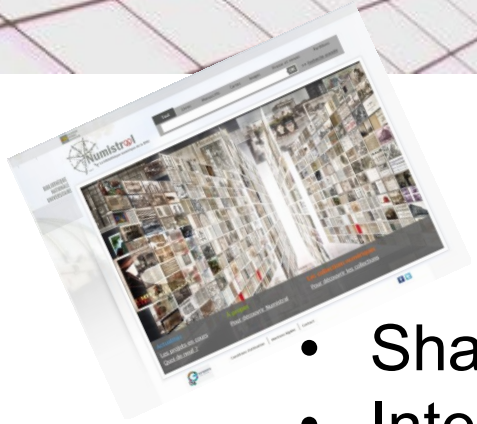
Preservation as a core process

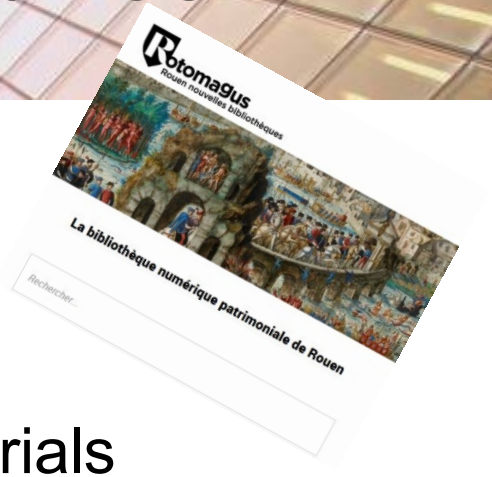


Agenda

- Context of digital assets at BnF
- From archiving to preserving
- Crafting a new organization
- Sharing data and workflows
- Handling the volume
- Conclusion

Collaborating with other libraries





- Sharing digitization
- Integrating already digitized materials
- Preserving the digital materials on their behalf
- Sharing the dissemination with Gallica
- Allowing for a specific design: ‘White mark Gallica’

« BnF – Archivage Numérique » offer

- **Arrangements are planned for integration,** monitoring and retrieval of the data:
  - ✓ Custom made by BnF team
  - ✓ Or by the « Espace Coopération» extranet of the BnF
- **Media provided** depending on the need:
  - On 2 tapes: access can take several hours
  - On 2 tapes and 1 disk: access in minutes, availability during working hours
  - On 2 tapes and 2 disks: access in minutes, availability 7d/7.
- **A standard contract** defines the Service Level Agreements (volume, duration, formats, metadata, rules, …)

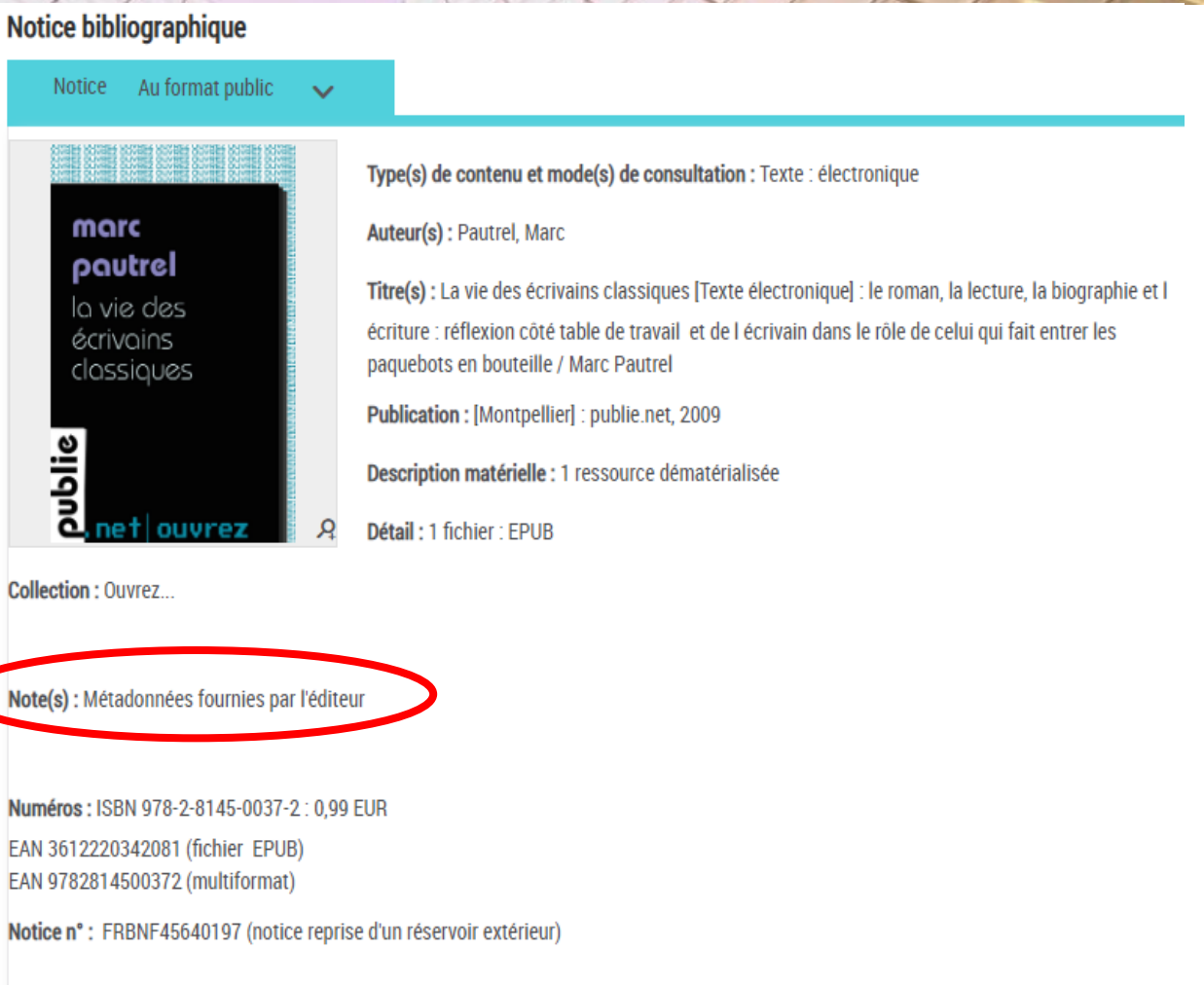
Third-party archiving costs

- It is a participation, at the best cost, in a spirit of sharing.
- Standard price lists are based on:
  - The way of downloading the data to archive (custom or extranet)
  - The kind of media (tapes, disks)
  - The volume to store (from 1 to 50 TB)
  - The duration of the initial contract (3, 5 or 8 years)

- The cost is derived from the service level of agreement.

Scalability in metadata acquisition

- Direct delivery from the publishers with dedicated B2B workflows



Notice bibliographique

Notice Au format public

Type(s) de contenu et mode(s) de consultation : Texte : électronique

Auteur(s) : Pautrel, Marc

Titre(s) : La vie des écrivains classiques [Texte électronique] : le roman, la lecture, la biographie et l écriture : réflexion côté table de travail et de l écrivain dans le rôle de celui qui fait entrer les paquebots en bouteille / Marc Pautrel

Publication : [Montpellier] : publie.net, 2009

Description matérielle : 1 ressource dématérialisée

Détail : 1 fichier : EPUB

Collection : Ouvrez...

Note(s) : Métadonnées fournies par l'éditeur

Numéros : ISBN 978-2-8145-0037-2 : 0,99 EUR

EAN 3612220342081 (fichier EPUB)
EAN 9782814500372 (multiformat)

Notice n° : FRBNF45640197 (notice reprise d'un réservoir extérieur)

Agenda

- Context of digital assets at BnF
- From archiving to preserving
- Crafting a new organization
- Sharing data and workflows
- Handling the volume
- Conclusion

Economical aspect: volume

Initial digitization
B&W - 300dpi -TIFF G4
1 page ~ **200KB**

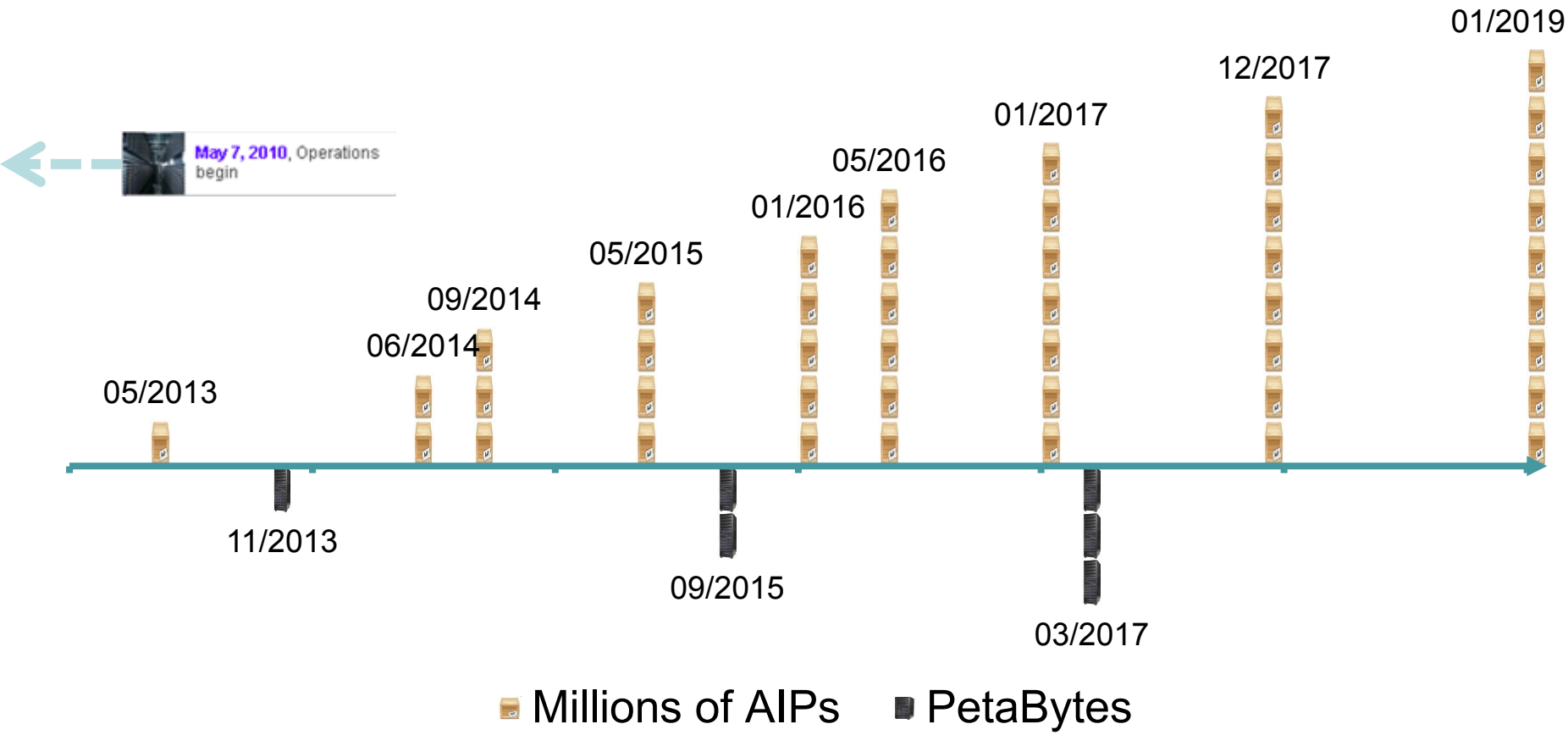
Current digitization
Color (24 bits) – 400dpi – uncompressed TIFF
1 page ~ **80MB**

Factor x400 !!!

Current digitization
Color (24 bits) – 400dpi – **JPEG 2000**
1 page ~ **20MB**

Divided by 4

Timeline of the growth



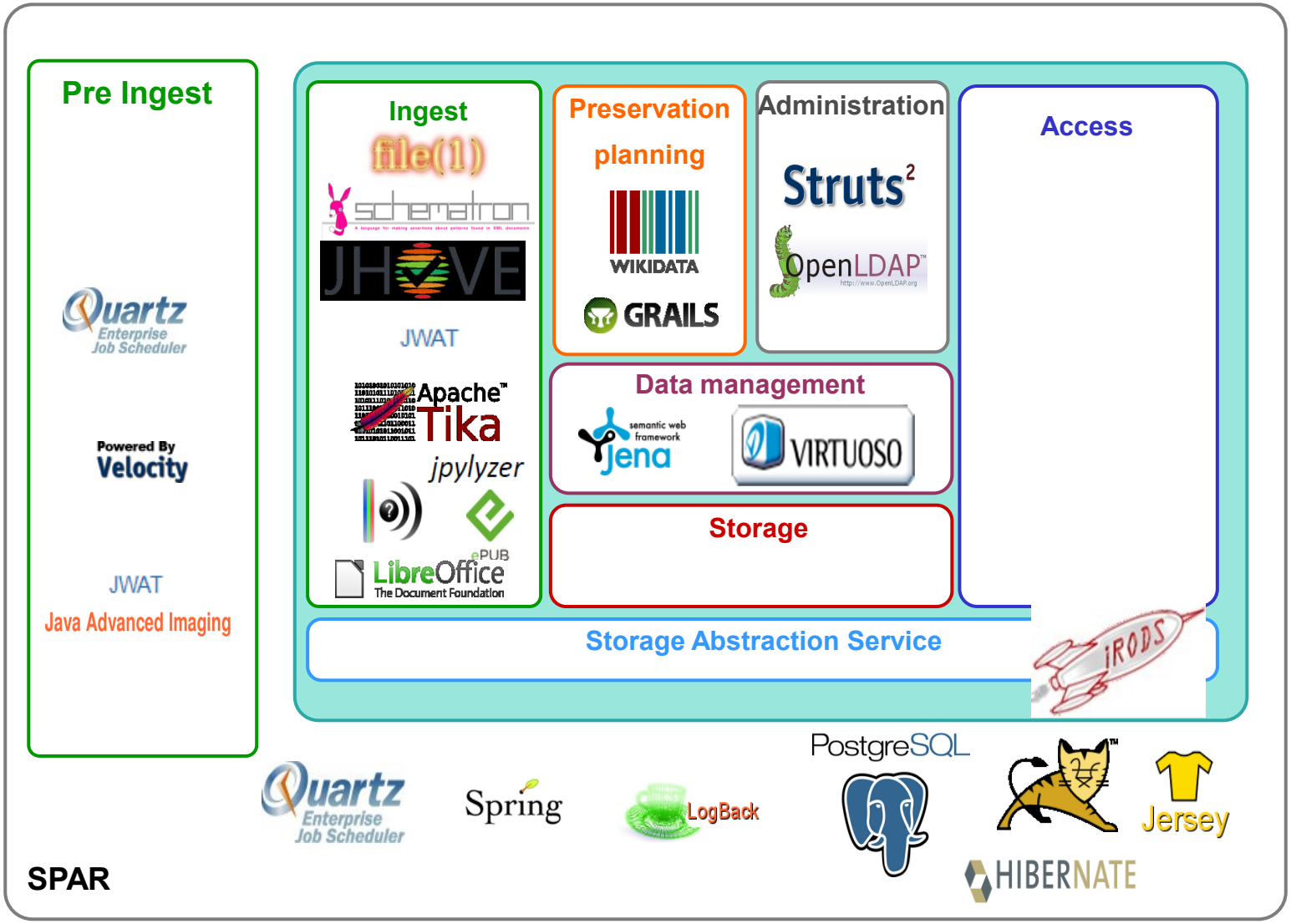
Economical aspect: storage infrastructure

Cloud still not an option: legal issues, recurrent cost, data transfer issue when moving from one provider to another (escape plan)

Hard-disk only for dissemination

Tapes remain economically viable (10 times less costly than hard-disk) and power efficient

Leveraging Open Source community effort



Agenda

- Context of digital assets at BnF
- From archiving to preserving
- Crafting a new organization
- Sharing data and workflows
- Handling the volume
- Conclusion

Conclusions

For the digital preservation to be a sustainable activity, it needs:

- Involvement of the institution (core mission, skills, training, …)
- Sharing with the ecosystem
- Building a business plan
- Always, anticipate risks

Thanks for your attention

Questions?

thomas.ledoux_at_bnf.fr

BnF