IFLA International Newspaper Conference


11-13 April 2012
BnF, Paris

With the support of:
Long-term preservation at BnF

Scalable Preservation and Archiving Repository (SPAR)
Agenda

• Context of digital preservation at BnF
• Overall schedule of the project
• Overall architecture and main features
• Service Level Agreements: how to build on mutual trust
• Conclusion
Technical issue: deal with the volume

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

Today digitization
Color (24 bits) – 400dpi
– TIFF uncompressed
1 page ~ 80Mb

More than x500 !!!
Functional issue: make a digital preservation copy

Digitization as a mean for preservation and dissemination
Business issue: native digital document

More and more documents exist only in digital form.

This is the form for the legal deposit and then for preservation.
Initial goals of the preservation

• Take the responsibility of digital data preservation
• Provide a storage facility for archives
• Reach a minimal critical mass to reduce costs (hardware, software and human)
• Allow for the pooling of archive among several institutions
How to manage digital archives

• Build a central repository to diminish the diversity (media, formats, departments …)
• To take inspiration from good practices and standards

⇒ Key requirements:
  – OAIS compliance
  – modularity and distributivity
  – abstraction
  – use of well known formats and standards
  – use of open-source technical building blocks
Digital archiving at BnF

Preservation digitization
Audiovisual
WEB Archiving

Production applications

Dissemination applications

Preservation planning
Data management
Administration
Access
Storage

Storage Abstraction Service (SAS)

SPAR - Realization
SPAR - Infrastructure

2012/04/13
Decomposition in channels

• To deal with the variability and heterogeneity of the data, definition of **channels**

• build on the relation between the digital objects and the archival system, independently of any given organization:
  – Preservation digitization
  – Audiovisual material
  – Negotiated legal deposit (dark Web, regional press)
  – Automatic legal deposit (surface Web)
  – Administrative production
  – Deposit / Third party archiving
  – Acquisition / Donation
Agenda

• Context of digital preservation at BnF

• Overall schedule of the project

• Overall architecture and main features

• Service Level Agreements: how to build on mutual trust

• Conclusion
Schedule

• Requirements definition (Working Groups) 2006
• Technical study of existing solutions 2006
• Implementation RFP 2007
• Iterative development 2008 - 2009
• Validation starts december 2009
• Production starts march 2010
Agenda

• Context of digital preservation at BnF
• Overall schedule of the project
• Overall architecture and main features
• Service Level Agreements: how to build on mutual trust
• Conclusion
Infrastructure

- Primary storage
- Secondary storage
- Lookup storage
- Backup lookup storage
- Backup secondary storage
- Backup storage
- Backup servers

Main site: (image)
Backup site: (image)
SPAR
a generic repository solution at BnF
The levels of granularity

set

Le Matin

Year 1882

01/07/1882 02/07/1882 03/07/1882

Year 1883

28/02/1883 01/03/1883

group

object

file

Slide 17

2012/04/13
Highlights from our information model: the seven deadly sections of METS

- Header
- **DmdSec**
- **AmdSec**
  - TechMD
  - DigiProvMD
  - SourceMD
  - RightsMD
- **FileSec**
- **StructMap**
- **Structlink**
- **BehaviorSec**

- Descriptive and source metadata: qualified Dublin Core
- Technical metadata: depends on the data-objects
- Provenance metadata: PREMIS
- Structural metadata: METS
- Not used in SPAR
Storage abstraction

SPAR - Realization

Main storage

Refreshment of media

Multiple copies management

Media migration

Backup storage

Application viewpoint

Infrastructure viewpoint
Agenda

• Context of digital preservation at BnF
• Overall schedule of the project
• Overall architecture and main features
• Service Level Agreements: how to build on mutual trust
• Conclusion
The Service Level Agreements

• The Service Level Agreements: contractualize with the users to offer a more transparent system

• The system is no longer a black box only known by technical experts
Taken into account in the system

- Which formats are allowed?
- What is the maximum size of a package?
- How copies are needed, in what kind of media?
- Do we need to log each access?

Storage abstraction service

SIP
AIP
DIP
 rdf
 mets
SLA: definition

- 3 SLA: Ingest, Preservation, Access
- Formalize in XML the ways of managing the packages
- Those 3 SLA are recorded in a reference package that describes the channel

Mets.xml
Contract.pdf
SLA-I.xml, SLA-P.xml, SLA-A.xml
Agenda

- Context of digital preservation at BnF
- Overall schedule of the project
- Overall architecture and main features
- Service Level Agreements: how to build on mutual trust
- Conclusion
Overview of an ingest
Information about a package

Nom : arkl2148-btv1b9054854m.version0.release0.tar

Catégorie : AIP

Statut : OK

MDS : 74a8b5e689462ba56d7ccd5524ec29de

Informations de stockage :

<table>
<thead>
<tr>
<th>Copie</th>
<th>Elément de stockage</th>
<th>Chemin</th>
<th>volume ID</th>
</tr>
</thead>
</table>
| copie 1 | capsAIP01         | /sas/elemAIP01-1/AIP/2792  
/2188/2294/ark-12148-btv1b9054854m.version0.release0.tar | a01706 74a8b5e689 |
| copie 2 | elemAIP01-2       | /sas/elemAIP01-2/AIP/2792  
/2188/2294/ark-12148-btv1b9054854m.version0.release0.tar | b00024 74a8b5e689 |

Processus associés au paquet :

Afficher toutes les actions

3 items trouvé(e)s, affichage des items.

<table>
<thead>
<tr>
<th>Jeton</th>
<th>Titre</th>
<th>Statut</th>
<th>Date de début</th>
<th>Date de fin</th>
</tr>
</thead>
<tbody>
<tr>
<td>a91ae50-62ad-1180-b49e-00144feac0aa</td>
<td>ING_1</td>
<td>completed</td>
<td>2011/04/09 15:31:22</td>
<td>2011/04/09 15:39:</td>
</tr>
<tr>
<td>b832bf80-62ae-11e0-a473-00144feac0aa</td>
<td>STD_1</td>
<td>completed</td>
<td>2011/04/09 15:38:57</td>
<td>2011/04/09 15:39:</td>
</tr>
<tr>
<td>caf2a70-62ae-11e0-8dfb-00144feac0aa</td>
<td>DM_2</td>
<td>completed</td>
<td>2011/04/09 15:39:29</td>
<td>2011/04/09 15:39:</td>
</tr>
</tbody>
</table>
Thanks for your attention

Questions?

Thomas Ledoux
thomas.ledoux_AT_bnf.fr