



**Archivage et  
préservation numérique  
à la  
Bibliothèque nationale de France**

**8 mars 2022**

# Ordre du jour

- BnF, bref historique et missions
- Préservation numérique à la BnF
- Modèle de données et Paquet de préservation
- Élaboration d'une politique de préservation
- Dépôt légal du né-numérique
- Coopération sur les bibliothèques numériques
- Tiers-Archivage

# Bref historique

- |                 |   |
|-----------------|---|
| 1368            | Installation de la première Bibliothèque royale au Louvre   |
| 1537            | édit de Montpellier, naissance du dépôt légal   |
| 1792            | la bibliothèque royale devient Bibliothèque nationale   |
| 14 juillet 1988 | le président François Mitterrand annonce la création d'une « bibliothèque d'un genre entièrement nouveau », tournée vers le numérique |
| janvier 1994    | création de l'établissement public national à caractère administratif, placé sous la tutelle du ministre de la culture                |

# La Bibliothèque nationale de France



7 sites

Personnel : 2 200 ETP

Budget : environ. 230 M€

Des missions locales, nationales et internationales

Environ 1M de lecteurs par an  
300 000 visiteurs aux expositions



Sites Web : 29 millions de visiteurs  
dont 15 millions pour Gallica



# De multiples missions

## Missions

- collecter le patrimoine national,
- le cataloguer et le préserver,
- en assurer un accès le plus large (localement et à distance),
- coopérer au sein de son réseau
- conduire des programmes de recherche

## Dépôt légal

- Dépôt légal imprimés depuis 1537
- 1648 : gravures et plans
- 1793 : enregistrements sonores
- 1925 : photogrammes
- 1938 : phonogrammes
- 1975 : vidéogrammes
- 1992 : documents électroniques
- 2006 : dépôt légal du Web
- 2021 : dépôt légal né-numérique

# Une collection diversifiée

## Collections

- + de 14M de livres et d'imprimés
  - + de 30M d'estampes et de photographies
  - + de 250 000 manuscrits
  - + de 900 000 documents sonores, 50 000 documents multimédia
- le dépôt légal de l'Internet français (milliards de fichiers)  
et aussi des partitions de musique, des médailles et objets antiques, des cartes, plans et globes géographiques, des objets de théâtre...

# Enjeu institutionnel : Missions

- La conservation est au cœur des missions de la BnF
- La conservation numérique se situe dans le prolongement direct de la conservation des collections de la BnF. Deux axes sont à considérer :
  1. le numérique comme moyen de conserver un document analogique qui se dégrade. Il y a alors **numérisation** (cas du papier mais aussi des collections audiovisuels analogiques) du document et besoin de conserver cet exemplaire numérique au même titre que l'exemplaire analogique (auquel il peut se substituer)
  2. le **document né numérique** (cas du dépôt légal du Web du 1<sup>er</sup> août 2006 ou du dépôt légal des documents numériques du 30/12/2021). Ce document entré dans les collections patrimoniales doit être conservé.



Décret n°94-3 du 3 janvier 1994 - article 2 :

La Bibliothèque nationale de France a pour missions de collecter, cataloguer, **conserver** et enrichir dans tous les champs de la connaissance, **le patrimoine national dont elle a la garde**, en particulier le patrimoine de langue française ou relatif à la civilisation française

# Ordre du jour

- BnF, bref historique et missions
- Préservation numérique à la BnF
- Modèle de données et Paquet de préservation
- Élaboration d'une politique de préservation
- Dépôt légal du né-numérique
- Coopération sur les bibliothèques numériques
- Tiers-Archivage

# Le projet SPAR : contexte

Applications de production de données



Numérisation de conservation

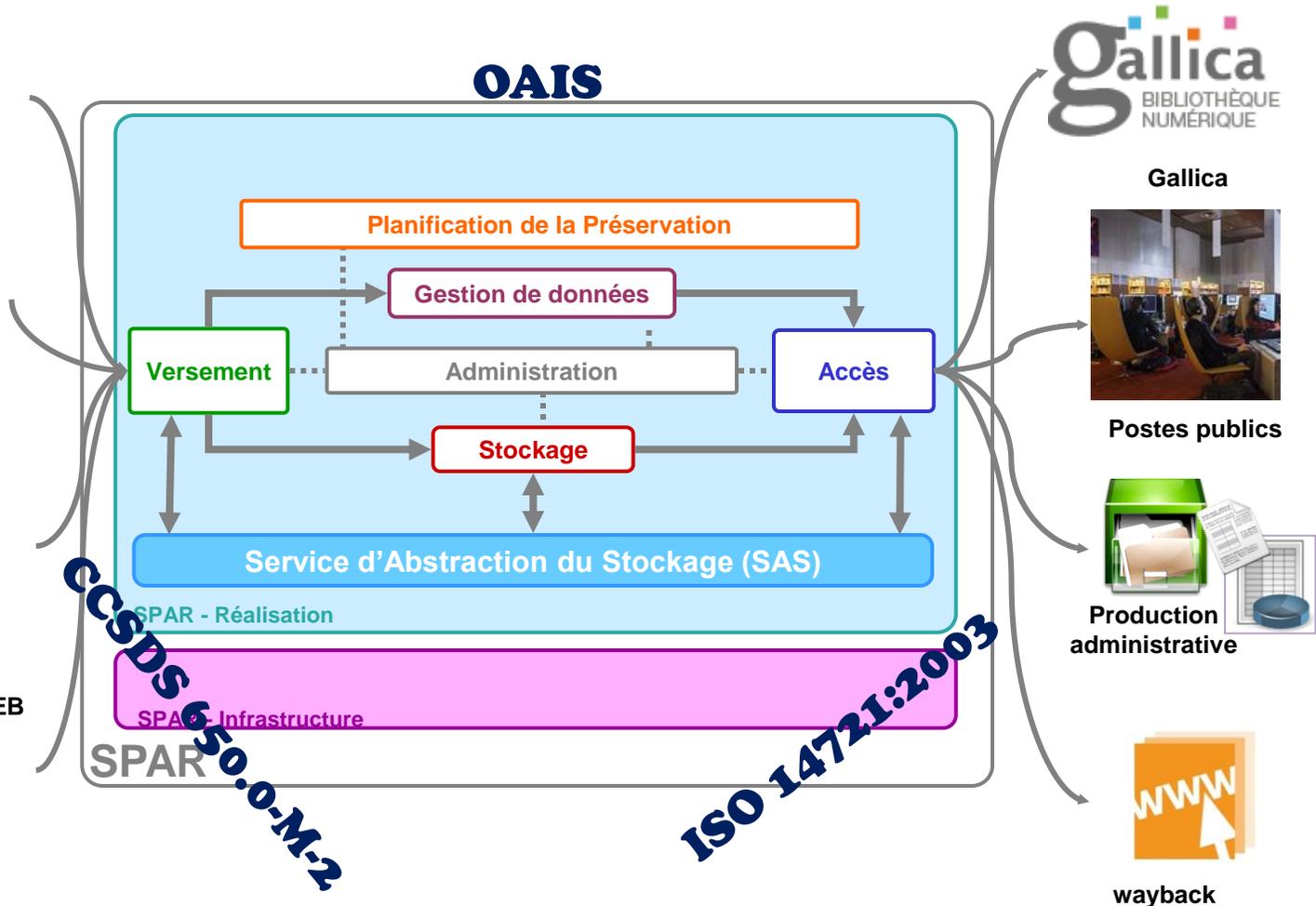


Production administrative  
Record Management



Archives du WEB

...



Applications de diffusion de données



Gallica



Postes publics



Production administrative

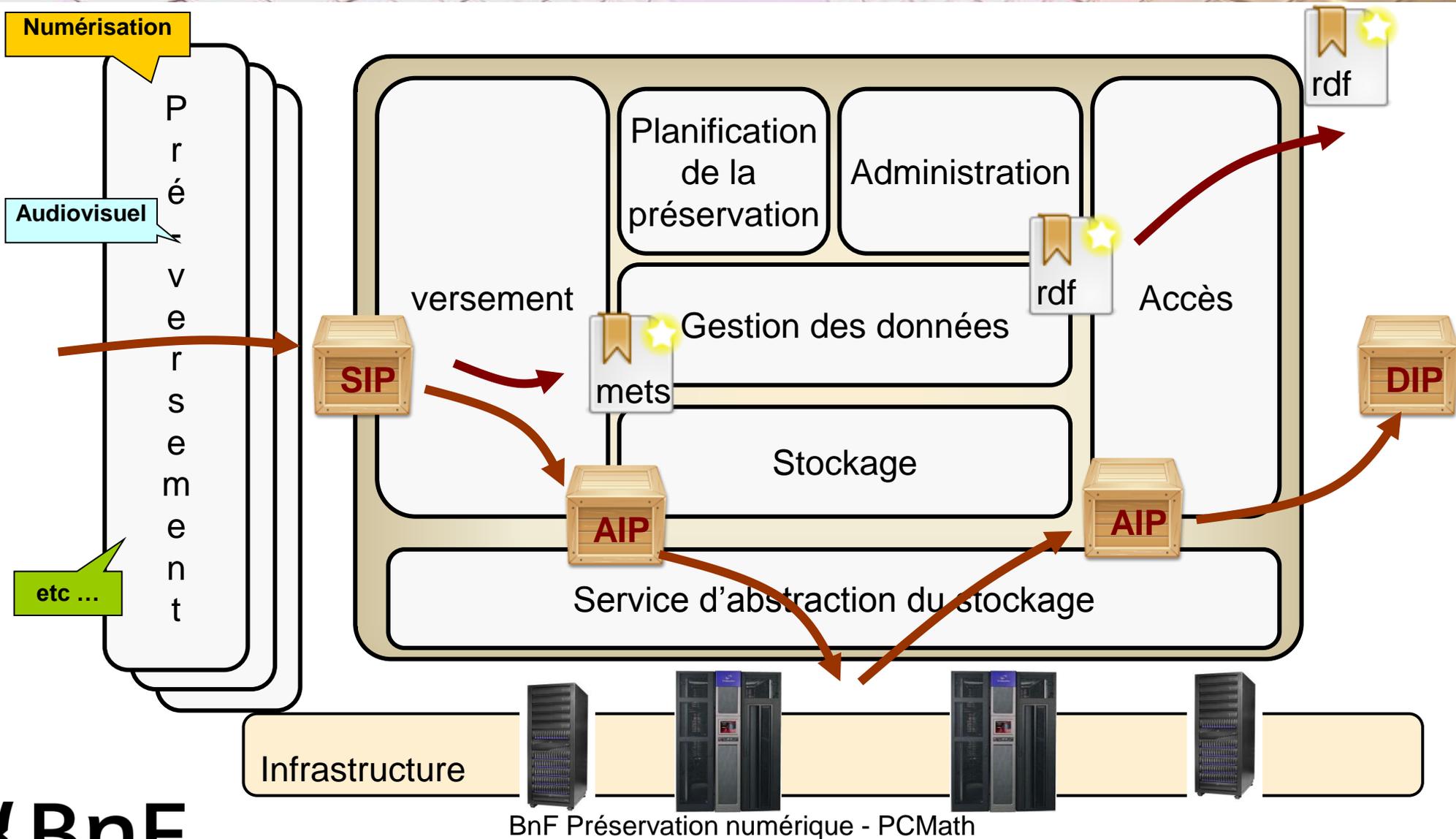


wayback

# SPAR en chiffres (au 03/03/2022)

- Ouverture de l'exploitation : mai 2010
  - Préservation de la numérisation courante de la BnF
  - Stockage répliqué sur deux sites
- 13 390 865 paquets archivés
- 495 720 882 objet-données (fichiers)
- > 4,9 Po (= 4 968 To) de données brutes
- > 7 milliards de triples (métadonnées élémentaires)
- Des manuscrits très volumineux ont été versés (paquets de taille supérieure à 100 Go)
- Deux migrations technologiques déjà réalisées, une nouvelle en préparation : nouvelle génération de bande
- Filières disponibles : monographies, périodiques, lots d'image, audio, vidéo, archivage du web, documents bureautiques, objets 3D, tiers archivage

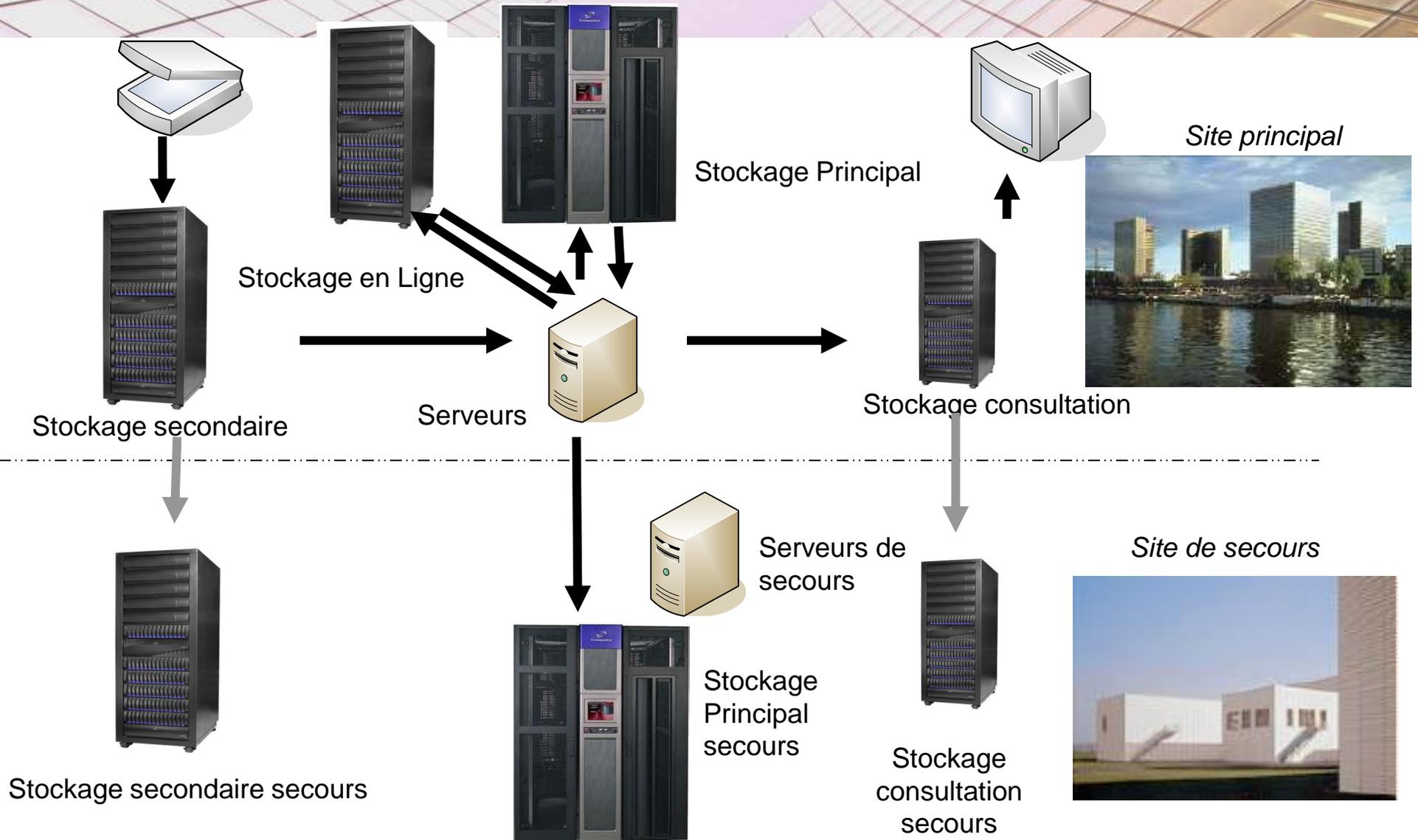
# Architecture générale



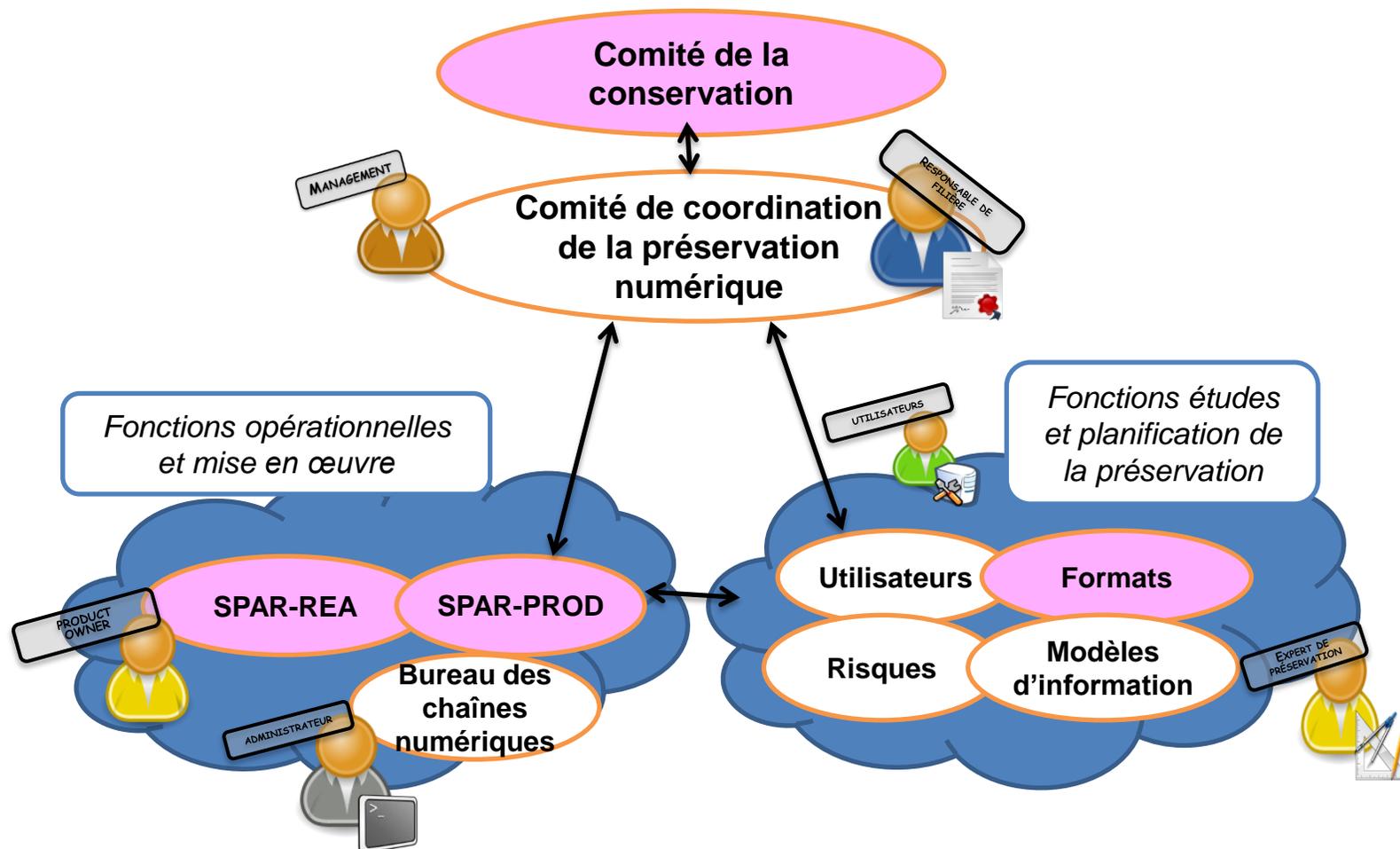
# Réalisation modulaire

- 7 modules
  - Module Versement
  - Module Accès
  - Module Administration
  - Module Planification de la préservation
  - Module Gestion des données
  - Module Stockage
  - Module technique « Service d'Abstraction du Stockage »
- Développement spécifique pour chaque chaîne
  - Pré versement (construction des SIP)
  - Post diffusion (reconstruction des paquets producteurs)

# Infrastructure matérielle



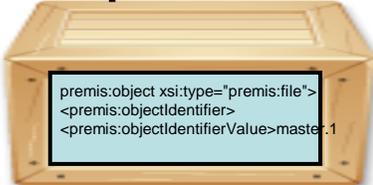
# Organisation de la préservation

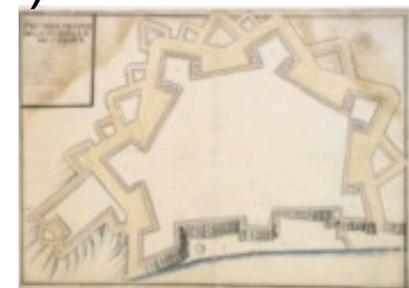


# Ordre du jour

- BnF, bref historique et missions
- Préservation numérique à la BnF
- Modèle de données et Paquet de préservation
- Élaboration d'une politique de préservation
- Dépôt légal du né-numérique
- Coopération sur les bibliothèques numériques
- Tiers-Archivage

# Que préserve-t-on?

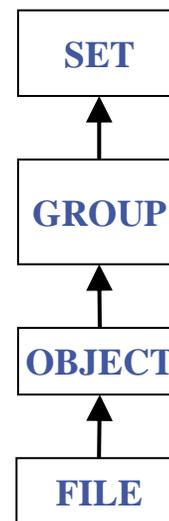
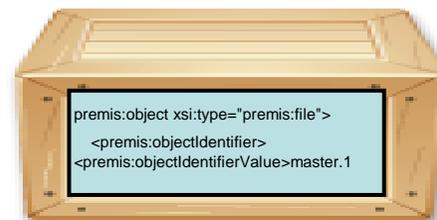
- Un document numérisé ou né numérique c'est :
  - Une description (métadonnées) : titre, auteur, date...
  - Des fichiers texte, image, ... 
  - Une structure physique (les fichiers sont ordonnés)
  - Une structure numérique
  - Les formats des fichiers
  - Les opérations effectuées...
- Tout cela dans un paquet complet et autonome qui est une enveloppe structurée englobant des données structurées ou non structurées 
- La même approche s'applique pour un document d'archive, pour un film, pour un livre



```
premis:object xsi:type="premis:file">  
<premis:objectIdentifier>  
<premis:objectIdentifierValue>master.1
```

# Les paquets d'archive (AIP)

- Une archive TAR contenant :
  - Un manifeste au format *METS* (métadonnées)
  - Ensemble des fichiers à préserver (données)
- Le paquet d'archive est un **paquet complet et autonome**
- Plusieurs niveaux de granularité :
  - set → **group** → object → file
  - ex: périodique → fascicule → page → image
- Liens entre paquets d'archive
  - lien structurel
  - liens logiques : collections, référence, documentation



# Les niveaux de granularité



Année 1882



Année 1883

set

Le Matin

group

01/07/1882

02/07/1882

03/07/1882

28/02/1883

01/03/1883

object



file

# Les schémas de métadonnées utilisés dans SPAR

Types d'information	Schémas de métadonnées
Informations techniques ou de représentation (format de données, caractéristiques des fichiers)	
Informations d'intégrité (taille du fichier, empreintes, état des support de stockage)	
Informations de provenance	
Information de contexte et d'environnement matériel et logiciel	
Informations descriptives ou bibliographiques (titres, auteurs, dates, sujets...)	
Informations de structure (arborescence ou carte des différents niveaux de granularité et des relations)	
Informations sur la structure numérique (liste des fichiers)	
Information de référence (identifiants, URI...)	

# Ordre du jour

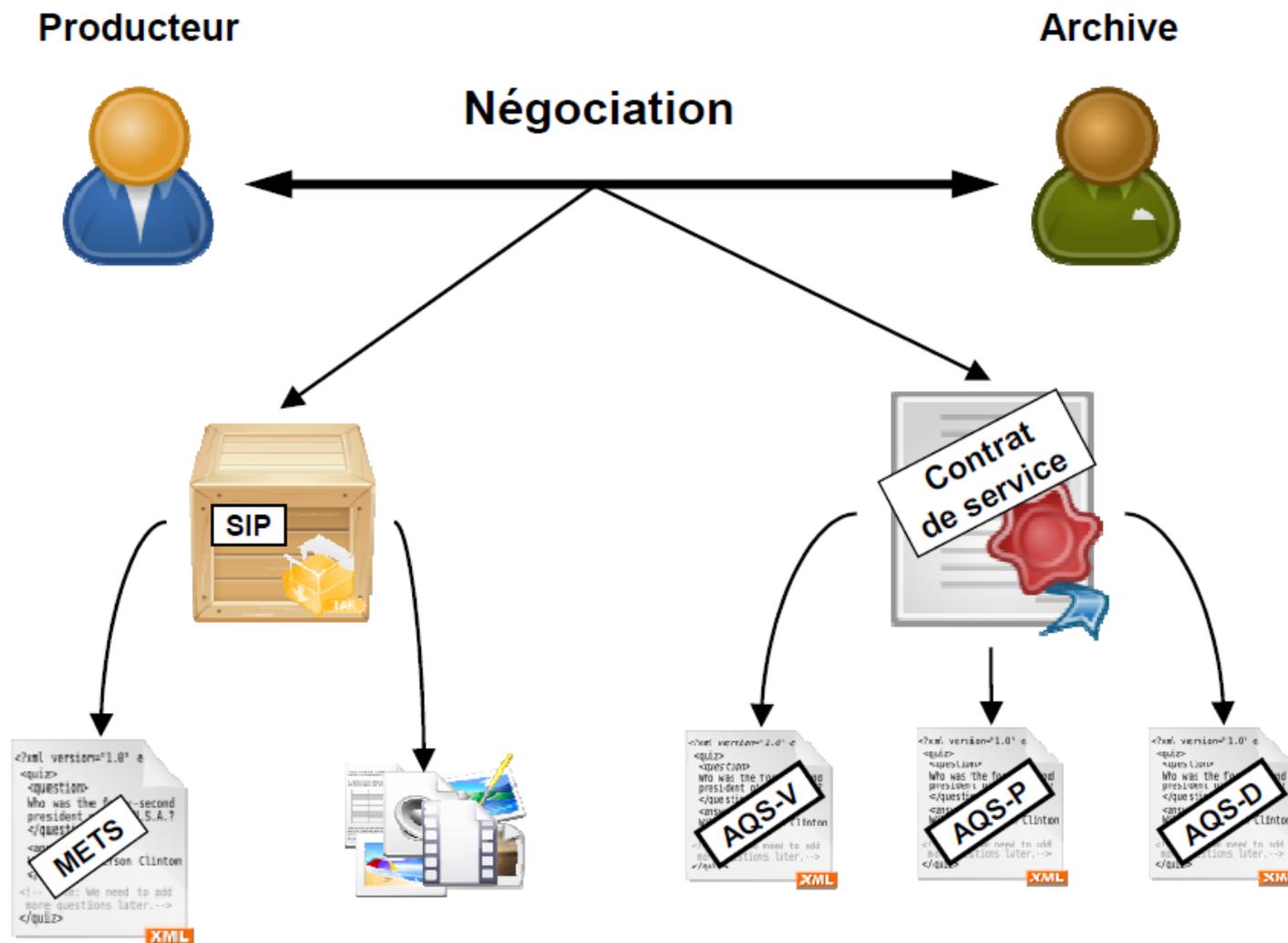
- BnF, bref historique et missions
- Préservation numérique à la BnF
- Modèle de données et Paquet de préservation
- **Élaboration d'une politique de préservation**
- Dépôt légal du né-numérique
- Coopération sur les bibliothèques numériques
- Tiers-Archivage

# La filière

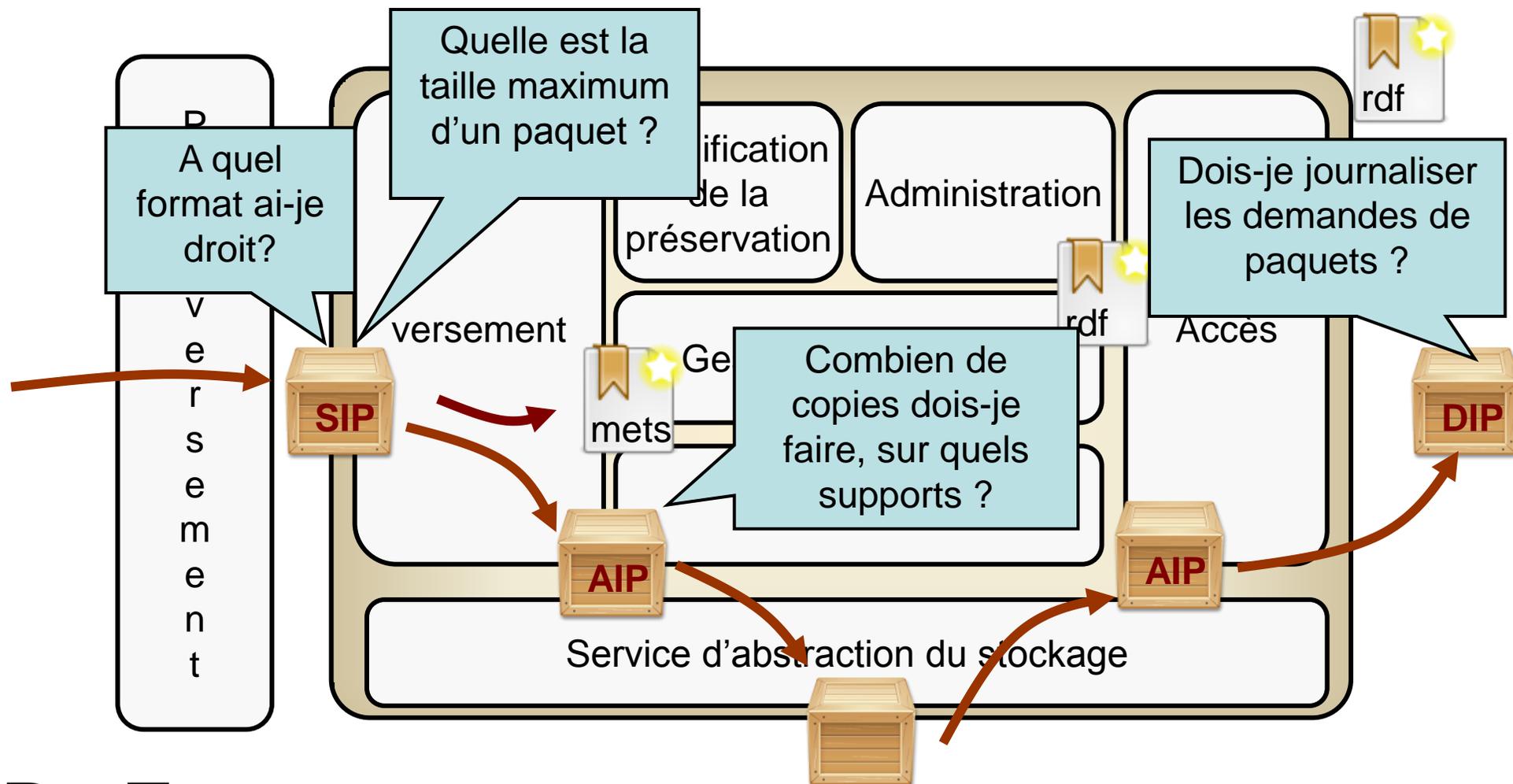
- Nécessaire pour gérer l'hétérogénéité des données ; spécialisée en chaîne
- Définie par les accords de qualité de service (AQS) entre les données et les exigences d'archivage :
  - Numérisation de conservation
  - Audiovisuel et multimédia
  - Numérisation de consultation
  - Dépôt légal automatique
  - Dépôt légal négocié
  - Production administrative et technique
  - Dons et acquisitions
  - Tiers archivage

**+ filière de référence**

# Principe : contractualiser la filière



# Accords de Qualité de service



# Exemple d'AQS de la chaîne Tiers Archivage

- Autoriser les formats suivants : tout ce qui est en xml ou en jpeg, et tous les mime types
- Calendrier : du lundi au vendredi (8h-20h)
- Taille maximale d'un paquet de 40Go
- 400 fichiers au maximum par paquet
- 2 000 paquets au maximum par jour
- Durée d'indisponibilité maximale de 48 heures
- Détection d'un fichier de métadonnée « manuscrit.xml », métadonnées à transformer selon les règles spécifiées dans le fichier « toDC.xsl »
  
- Une copie disque + une copie bande sur site principal et une copie bande sur le site de secours
- Un audit tous les six mois

# AQS: définition

- 3 AQS: Versement, Préservation, Accès
- Formalise en xml les conditions d'exploitation des paquets
- Ces 3 AQS sont enregistrés dans un paquet de référence descriptif de la chaîne



Mets.xml



Contrat de service.pdf

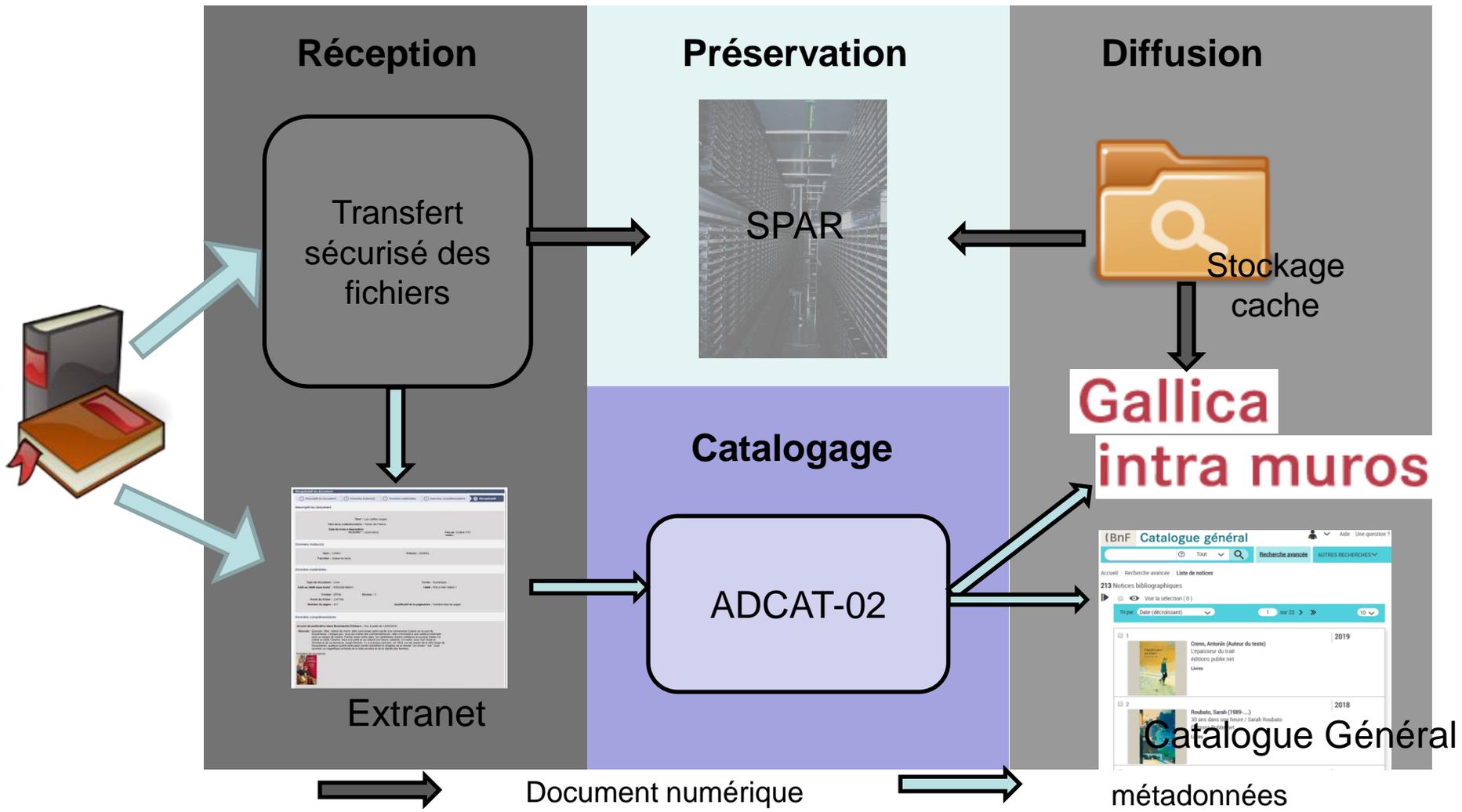


AQS-V.xml, AQS-P.xml, AQS-D.xml

# Ordre du jour

- BnF, bref historique et missions
- Préservation numérique à la BnF
- Modèle de données et Paquet de préservation
- Élaboration d'une politique de préservation
- Dépôt légal du né-numérique
- Coopération sur les bibliothèques numériques
- Tiers-Archivage

# Dépôt légal du son dématérialisé



# Catalogage / Signalement

## Notice bibliographique

Notice Au format public



**Type(s) de contenu et mode(s) de consultation :** Musique exécutée : audio ou électronique

**Interprète(s) :** [Perret, Guillaume \(1980-....\)](#). Interprète

**Titre(s) :** A Certain Trip / Guillaume Perret

**Diffusion/distribution :** [France] : Idol, 2020 (diffusion/distribution)

**Description matérielle :** 1 ressource dématérialisée (2 fichier(s) audio : 000944)

**Détail :** 2 fichiers audio : FLAC (44,1 kHz)

**Note(s) :** **Métadonnées fournies par l'éditeur**

Copyright : 2020 Antipodes Music Productions  
**Production phonographique :** Production phonographique : 2020 Antipodes Music Productions

**Genre :** jazz, blues et gospel > jazz

**Genre non contrôlé :** Jazz: Contemporary Jazz (éditeur)

**Circuit de distribution :** édition phonographique

**Numéros :** EAN 3663729113900

**Référence(s) commerciale(s) :** 397115 (Antipodes Music Productions)

**Outils**

Citer la notice : <https://catalogue.bnf.fr/ark:/12148/cb>

Télécharger/Imprimer

Envoyer par courriel

Ajouter à mes notices

Signaler une erreur sur cette notice

**LOCALISER CE DOCUMENT**  
(2 EXEMPLAIRES)

**Document numérique :**

**AVNUM-1**  
support : son dématérialisé

Consultable sur place en Bibliothèque de recherche

**Tolbiac - Rez-de-jardin - magasin - Salle P (type de place audiovisuelle (vidéo/son))**

**SNUM-2152**  
support : son dématérialisé

Réserver

<https://catalogue.bnf.fr/ark:/12148/cb46542455v>

La création initiale de la notice bibliographique se fait à l'aide des informations fournies directement par l'éditeur

# Le système de préservation comme un magasin numérique

- Avec les documents né-numériques, l'entrée des documents se fait d'abord dans le magasin numérique
- Cela est vrai pour le signalement comme pour le versement des données
- La mise à disposition est un processus postérieur :
  - Extraction de chaque nouveau paquet
  - Création des artefacts de consultation
  - Mise à disposition de ces derniers

# Diffusion alimentée directement par le magasin numérique

The image shows a screenshot of the BnF Gallica website interface. The top navigation bar includes the logo 'Gallica intra muros', a search bar, and options for 'RECHERCHE AVANCÉE', 'SÉLECTIONS', and 'FR'. The main content area is divided into two panels. The left panel, titled 'A Certain Trip / Guillaume Perret', shows a 'SYNTHÈSE' section with 'Enregistrements sonores' and 'SOUS DROITS' buttons. Below this are sections for 'EN SAVOIR PLUS', 'A DÉCOUVRIR', and 'Sujets similaires' with a 'jazz' category. The right panel, titled 'Le géant du ténor / John Coltrane', features a digital audio player with a progress bar and volume control. Below the player is a table of tracks:

Plage	Titre	Durée
Plage 1	Giant Steps	0:04:43
Plage 2	Cousin Mary	0:05:46
Plage 3	Countdown	0:02:22
Plage 4	Spiral	0:05:57

Below the table is a section for 'Auteurs similaires' featuring 'Coltrane, John' and three album covers. At the bottom of the right panel, a vinyl record cover for 'ATLANTIC "LE GEANT DU TENOR"' is displayed, listing tracks 1. GIANT STEPS and 2. COUSIN MARY.

Un disque “numérique natif” se présente de la même manière qu’un disque numérisé

# Donner accès

- Limité aux emprises de la Bibliothèque
- Utilisation d'un "Virtual Browser" pour ne donner qu'à voir l'information sans donner un accès direct aux fichiers

## Site François-Mitterrand

Où consulter un document dématérialisé sous dépôt légal ?

**Personnel BnF**  
Accès au dépôt légal numérique via Gallica intra muros et les Archives du web, **sans export possible**

**Bibliothèque d'étude (Haut-de-jardin)**  
**PAS D'ACCES AU DEPOT LEGAL NUMERIQUE**  
Que ce soit sur l'un des 309 postes publics ou sur un poste personnel relié au réseau filaire via le Wi-Fi

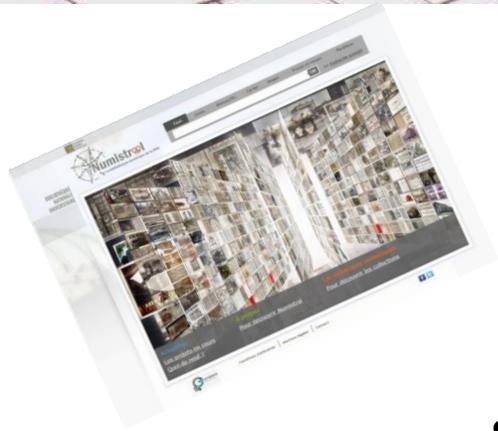
**Bibliothèque de recherche (Rez-de-jardin)**  
Contenus accessibles uniquement sur accréditation  
**ACCES AU DEPOT LEGAL NUMERIQUE**  
Sous certaines conditions  
Via Gallica intra muros, les postes audiovisuels dédiés et les Archives de l'Internet  
Accès complet sur l'un des 240 postes publics, **sans export possible**  
Accès limité\* sur l'un des 63 postes audiovisuels et multimédia (PAV), **sans export possible**  
Accès sur un poste personnel, limité\*\*, **sans export possible**, à la condition d'être relié au réseau filaire, ou relié au réseau Wi-Fi avec authentification du chercheur et contrôle de son accréditation

\* limité aux fichiers au format FLAC uniquement  
\*\* limité aux seuls livres numériques en 2018

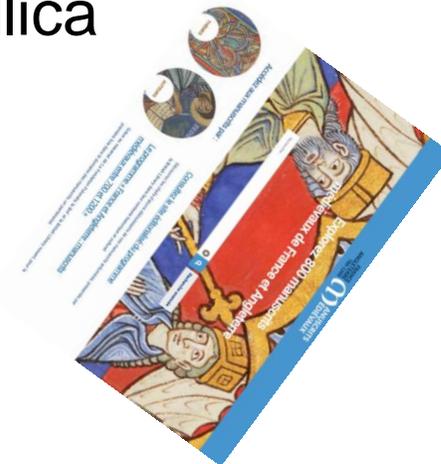
# Ordre du jour

- BnF, bref historique et missions
- Préservation numérique à la BnF
- Modèle de données et Paquet de préservation
- Élaboration d'une politique de préservation
- Dépôt légal du né-numérique
- Coopération sur les bibliothèques numériques
- Tiers-Archivage

# Collaboration avec d'autres bibliothèques



- Partager la numérisation
- Intégrer des fonds déjà numérisés
- Préserver les données numériques
  - Partager la diffusion par Gallica
- Proposer une interface en propre : « Gallica marque blanche »



# Exemple de périodique

{BnF Gallica

TOUT GALLICA

Rechercher...

RECHERCHE AVANÇÉE

TOUTES NOS SÉLECTIONS PAR TYPES DE DOCUMENTS PAR THÉMATIQUES PAR AIRES GÉOGRAPHIQUES BLOG

Accueil > Années disponibles

Bulletin des sciences mathématiques et astronomiques

Rechercher

Rechercher dans tous les numéros

Informations détaillées

Notice complète

**Titre :** Bulletin des sciences mathématiques et astronomiques  
**Éditeur :** (Paris)  
**Date d'édition :** 1870-1884  
**Contributeur :** Darboux, Gaston (1842-1917). Éditeur scientifique  
**Contributeur :** Hoüel, Jules (1823-1886). Éditeur scientifique  
**Contributeur :** Loewy, Maurice (1833-1907). Éditeur scientifique

10 années disponibles - 11 numéros

1870 1 numéro	1871 1 numéro	1872 1 numéro	1873 1 numéro	1874 1 numéro	1875 1 numéro	1876 2 numéros	1877 1 numéro	1878 1 numéro	1879 1 numéro
1880 1 numéro	1881 1 numéro	1882 1 numéro	1883 1 numéro						

<https://gallica.bnf.fr/ark:/12148/cb34444944h/date>

# Deux exemples de marque blanche

The screenshot displays the website for the Bibliothèque Numérique de la Statistique Publique (BNSP). At the top left is the BNSP logo and the text 'BIBLIOTHÈQUE NUMÉRIQUE DE LA STATISTIQUE PUBLIQUE'. To the right are links for 'AIDE / FAQ' and 'EXPLORER', along with a language selector set to 'Fr'. A 'MENU DÉTAILLÉ' icon is on the far left. The main header features the title 'Les publications de la statistique publique' and a search bar with the placeholder 'Rechercher...' and a 'Recherche avancée' button. Below the search bar is a navigation menu with 'Explorer par' and four categories: 'THÉMATIQUE' (with a 3x3 grid icon), 'RÉGION & PAYS' (with a location pin icon), 'PÉRIODE' (with a timeline icon), and 'PARTENAIRE' (with a line icon). The main content area includes a photograph of industrial buildings with a sign 'ATELIERS DE MONTEUIL JECLAIRE' and a text block: 'Explorez les 15 000 documents numérisés de l'École nationale des ponts et chaussées, un document dans l'ensemble des collections anciennes et des archives.' Below this is a section titled '— Les essentiels' with five featured items: 'Histoire de l'École et de ses collections', 'Tous les outils et instruments de recherche', 'Chercher un ingénieur des Ponts', 'Cours de l'École', and 'Les Annales des ponts et chaussées'. At the bottom right of the screenshot is the text 'Bibliothèque numérique de la statistique publique' and the URL <https://www.bnsp.insee.fr/bnsp/>.

L'Héritage des ponts et chaussées  
<https://heritage.ecoledesponts.fr/enpc/>

# Affichage d'un fascicule en PDF « natif »

**BNSP** BIBLIOTHEQUE NUMERIQUE DE LA STATISTIQUE PUBLIQUE

Rechercher... **RECHERCHE AVANCÉE**

Accueil > 23 036 résultats page 1 sur 1 536 > Consultation

## Les études de marché et sondages à l'heure du big data

Institut national de la statistique et des études économiques (France). Auteur du texte

**SYNTHÈSE** ▾

- Livres

**BNSP**

EN SAVOIR PLUS >

A DÉCOUVRIR >

### Les études de marché et sondages à l'heure du big data

ip1890-figures\_fr.pdf

↓ [Télécharger le PDF \(1 MB\)](#)

Comment lire un PDF ?



*Ce document est accessible uniquement en téléchargement*

[Afficher les éléments d'accompagnement](#)

ÉLÉMENT(S) D'ACCOMPAGNEMENT (1) | ↓ [Télécharger les éléments \(23 KB\)](#)

<https://www.bnsp.insee.fr/ark:/12148/bc6p079mz0r>

# Affichage d'éléments d'accompagnement



BIBLIOTHÈQUE NUMÉRIQUE  
DE LA STATISTIQUE PUBLIQUE

Rechercher...



RECHERCHE  
AVANCÉE

Accueil > 23 036 résultats page 1 sur 1 536 > Consultation

## Les études de marché et sondages à l'heure du big data

Institut national de la statistique et des études économiques (France). Auteur du texte

SYNTHÈSE >

EN SAVOIR PLUS >

**Titre :** Les études de marché et sondages à l'heure du big data

**Auteur :** Institut national de la statistique et des études économiques (France).

Auteur du texte

**Éditeur :** Insee (Paris)

**Date d'édition :** 2022

**Sujet :** Sondages d'opinion

**Sujet :** Instituts de sondages

**Périodique lié :** INSEE première(ISSN

1950-6252) (Notice :

<http://catalogue.bnf.fr/ark:/12148/cb4014>

**Notice du catalogue :**

<http://catalogue.bnf.fr/ark:/12148/cb4696>

**Relation :** Première de couverture :

648469

**Type :** monographie imprimée

**Type :** Document électronique

**Langue :** français

**Format :** application/pdf

**Format :** Nombre total de vues : 1

**Description :** Comprend : Autres auteurs :

A DÉCOUVRIR >



Les études de marché et sondages à l'heure du big data

↓ Télécharger le PDF (1 MB)

Masquer les éléments d'accompagnement

ÉLÉMENT(S) D'ACCOMPAGNEMENT (1)

↓ Télécharger les éléments (23 KB)



ip1890-  
figures\_fr\_IP1890.xlsx

Poids : 23 KB

<https://www.bnsf.insee.fr/ark:/12148/bc6p079mz0r>



# Ordre du jour

- BnF, bref historique et missions
- Préservation numérique à la BnF
- Modèle de données et Paquet de préservation
- Élaboration d'une politique de préservation
- Dépôt légal du né-numérique
- Coopération sur les bibliothèques numériques
- Tiers-Archivage

# Extranet Archivage Numérique

ARCHIVAGE  
NUMÉRIQUE

{BnF} Bibliothèque nationale de France

## {BnF} Archivage numérique

Suivi d'archivage

Utilisateurs

> Déconnexion

Votre chaîne Tiers Archivage est CNACGP

LISTE DES VERSEMENTS

RECHERCHER DES DOCUMENTS

SUIVRE LES DEMANDES DE RESTITUTIONS

SUIVI GLOBAL

SUIVI DÉTAILLÉ

106 résultats trouvés, éléments de 1 à 50.

Filtrer date versement:

Debut

Fin

FILTRES

EFFACER

<< < 1 2 3 > >>

<input type="checkbox"/>	DATE DE BILAN	DATE DU DERNIER VERSEMENT	NOMBRE DE PAQUETS VERSÉS	TAILLE TOTALE DES PAQUETS VERSÉS	NOMBRE D'OBJETS DANS LA CHAÎNE	NOMBRE MOYEN D'OBJETS PAR PAQUET
<input type="checkbox"/>	2021/10/01 01:00:17	2021/09/30 18:48:43	341 270	218,4 To	9 373 304	27
<input type="checkbox"/>	2021/09/06 17:28:00	2021/08/23 16:02:20	340 201	216,7 To	9 358 021	27
<input type="checkbox"/>	2021/08/01 01:00:22	2021/07/31 00:47:10	339 718	209,6 To	8 888 088	26
<input type="checkbox"/>	2021/07/01 01:00:21	2021/07/01 00:58:46	338 531	203,1 To	8 440 926	24
<input type="checkbox"/>	2021/06/01 01:00:22	2021/05/31 21:38:47	337 897	202,7 To	8 392 225	24
<input type="checkbox"/>	2021/05/01 01:00:20	2021/04/30 20:41:15	336 628	200,2 To	8 236 716	24
<input type="checkbox"/>	2021/04/01 01:18:39	2021/03/31 23:56:07	336 243	195,8 To	8 080 916	24
<input type="checkbox"/>	2021/03/01 01:16:01	2021/02/27 15:17:13	335 645	190,4 To	7 679 297	22
<input type="checkbox"/>	2021/02/01 01:15:47	2021/01/31 22:37:33	334 943	187,9 To	7 454 896	22
<input type="checkbox"/>	2021/01/01 01:00:17	2020/12/31 15:25:06	334 393	180,2 To	7 070 445	21

# Modalités



- Mutualisation du système de préservation et de l'expertise afférente
- Données à la main du tiers-archivé.
- Convention directement appliqué sur l'accord de qualité de service.
- Tarifs standards basés sur :
  - Le mode de téléchargement des données à archiver (extranet ou sur-mesure)
  - Le support de stockage (bandes, disques)
  - La volumétrie de stockage (de 1 à 50 To)
  - La durée du contrat initial (3, 5 ou 8 ans).



# Salle informatique principale

