



ARCHIVAGE NUMÉRIQUE

{BnF | Bibliothèque
nationale de France

Plan

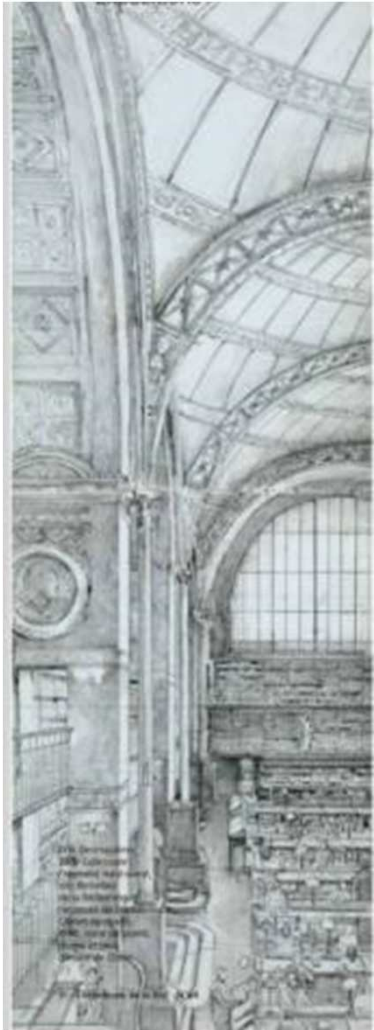
- La mission de conservation
- Les collections numériques à la BnF (action de préservation ?)
- Les services apportés par SPAR
- Les filières dans SPAR
- La filière tiers-archivage (agrément + tarifs)
- Pensez à récupérer le logo IMPACT

Enjeu de la conservation à la BnF



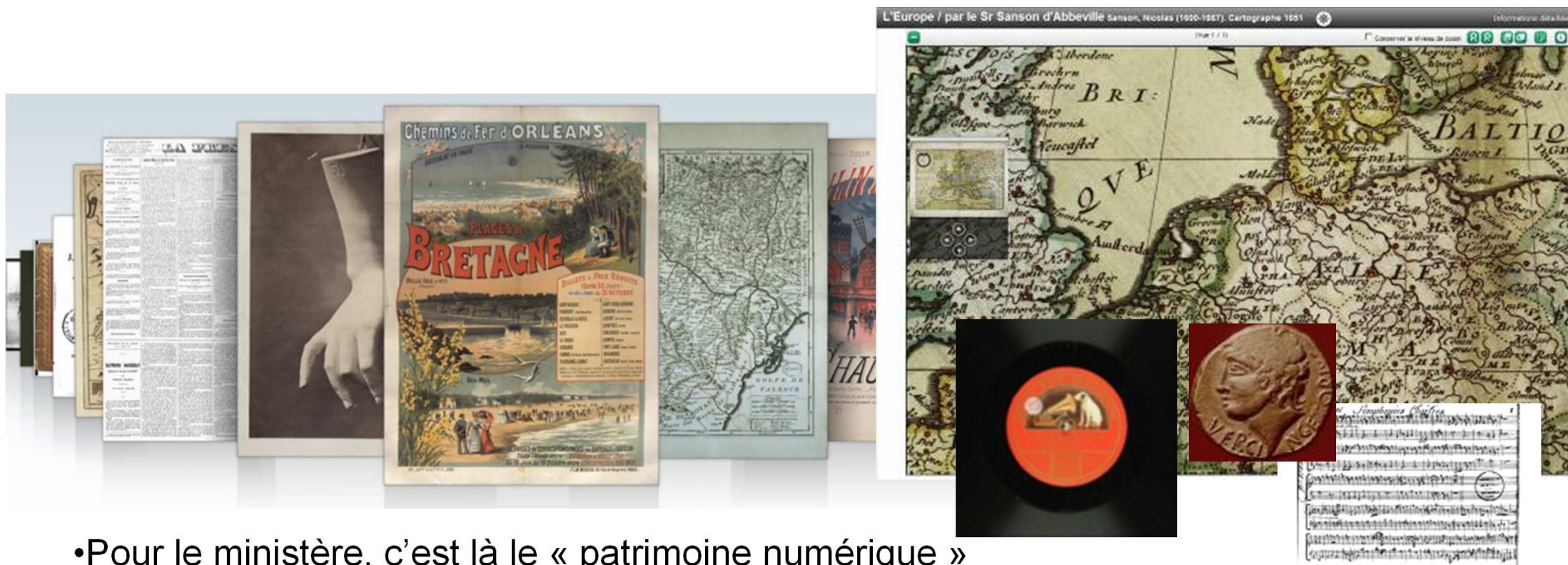
Décret n°94-3 du 3 janvier 1994 - article 2 :

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



- La conservation est au cœur des missions de la BnF : elle conserve, depuis des siècles, tout ce qui est publié en France.
- Constituées depuis 1537 notamment grâce au dépôt légal, ses collections représentent plus de 30 millions de livres, manuscrits, journaux, photographies, musiques, parmi lesquels un grand nombre de pièces uniques.
- 400 km de rayonnage

Du numérisé



- Pour le ministère, c'est là le « patrimoine numérique »
 - <http://www.numerique.culture.fr/pub-fr/index.html>
- Déjà plusieurs « strates » de numérisation ...

Qu'a-t-on numérisé ?

- Du papier : livres, images, presse, etc.
- Des objets : maquettes, statues, etc.
- De l'audiovisuel : cassettes audio, VHS, disquettes, CD, CD-ROM, etc.
 - qui nécessite un outil de lecture
 - qui est parfois déjà numérique
- Des originaux rares, précieux, fragiles
- Des versions numériques plus ou moins fidèles à l'original

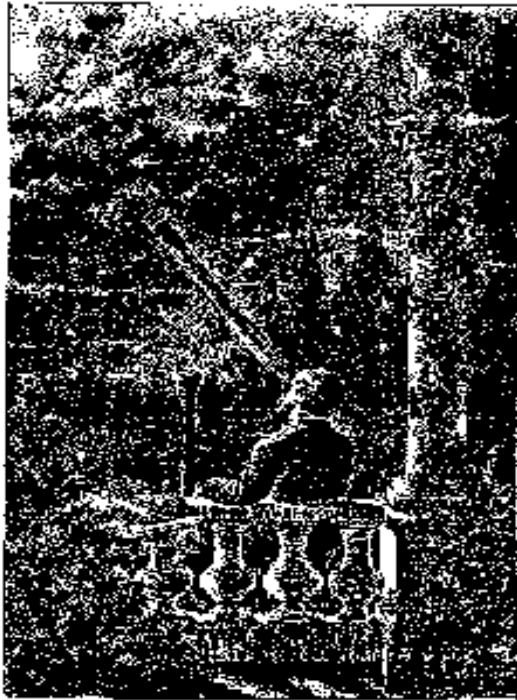
Dans quel cadre ?

- Quel est le statut des documents numériques produits ?
 - Hors droits, sous droits
 - Numérisés « à des fins de conservation » ou pour « préserver les conditions de sa consultation » au sens du Code de la propriété intellectuelle (article L122-5)
 - Dans le cadre d'un contrat passé avec un auteur
 - Dans le cadre d'une convention passée avec une bibliothèque partenaire

Comment ?

- Choix techniques :
 - Liés aux objectifs
 - Diffusion, préservation, les deux
 - Aux moyens techniques à une date donnée
 - Format de fichier, résolution, couleur
 - Mode image, texte (OCR), table des matières et autres enrichissements (entités nommées, etc.)

Comparons...



Débuts de la numérisation
N&B - 300dpi - TIFF G4
1 page ~ **200Ko**



Numérisation maintenant
Couleur– 400dpi – TIFF non
compressé
1 page ~ **80Mo**

Plus de 500 fois plus !

Du né-numérique

- Une masse et une diversité croissantes
- Beaucoup de questions



Livres et périodiques

- Acquis ou accessibles ?
- Sous licence locale ou globale ?
- Catalogués ou indexés ?
- Libres ou protégés (DRM) ?
- PDF, HTML, EPUB, indéterminés ?
- Quelles relations au « papier » : copie exacte, contenu similaire, genre totalement nouveau ?
 - Sans oublier les éditions électroniques de la bibliothèque elle-même...

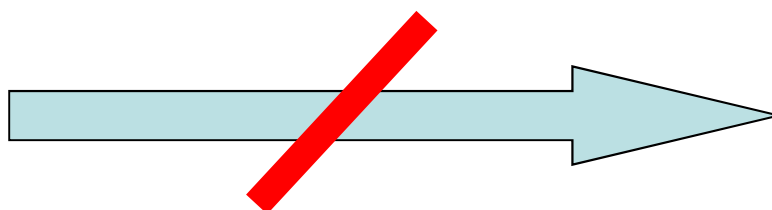
Comparons...

Le Monde



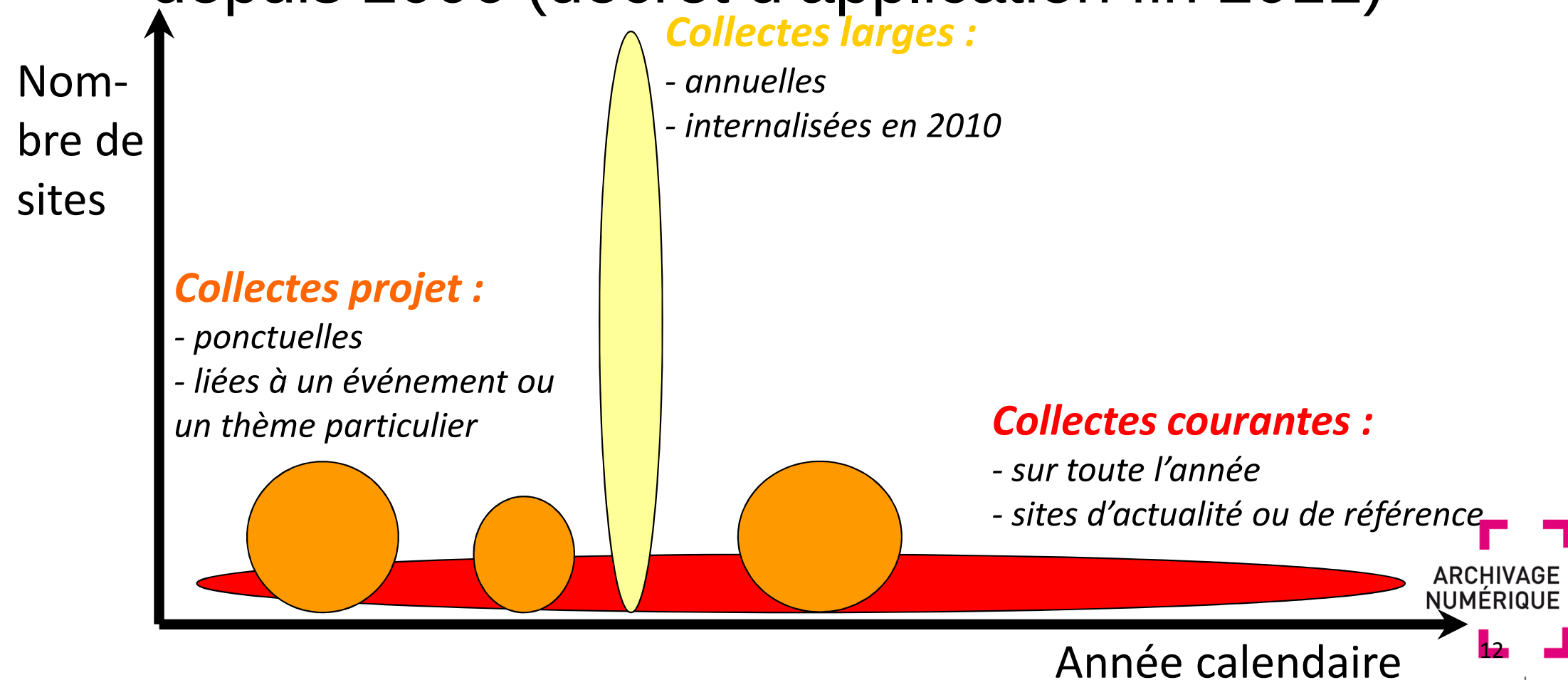
Le Monde.fr

France-soir



Internet

- Dépôt légal de l'internet : à la BnF et à l'INA, depuis 2006 (décret d'application fin 2011)



Archives

- Archives publiques
- Archives privées, achetées ou reçues
- Quels droits sur les documents ?
 - Conservation pendant la période d'utilisation courante (GED), et après ?
- Quelle variété de documents ?
 - Progiciels, environnements numériques de travail, e-mails, etc.

Les activités d'une bibliothèque

Sélection

Acquisition

Catalogage

Équipement

Magasinage

Conservation

Communication

Valorisation

Sélection

Collecte / production

Métadonnées / indexation

Empaquetage

Gestion

Préservation

Accès

Valorisation

0010000000100000001000000011000100110010001100000010000000100000
0010000000100000001101010011100000100000001000000011010100100000
0010000000100000001100010011100100111000001100010010110100110000
0011100100101101001100110011000001010100001100010011010000111010
0011001000110010001110101011000110001100011011000110010
0011100101011010001000000010000000100000001100110011100000111001
0011100000101110001101110011000000100000001000000010000000110010
0011001100110110010010110010001011000110100011010000000100000
001000000010000000100000001000000010000000110001100110
001000000010000000100000001011010011011001011100110111001100100
0100000001000000010110100110001001110000010111000110110001100010
010000000100000001000000010000000100000001000000010000000
1000000010000000101100111000001100000010000000100000001000000
1000000010000000100000001100010010000000100000001101000011001100
1000000010000000100000001110010010000000100000001000000010000000
1110010011100000111000000010000000100000001000000010000000
100000001000000010000000100000001000000010000000100000001001000
1000000010000000100000001000000010000000110101001100110010111000
1110000011001100100000001000000010000000100000001000000011010100
1101100010111000110110001000000010000000100000001000000010000000
1000000010000000110100001011100011100000110001001000000010000000
1000000010000000110111001110011001100110010111000110101001110000
10000000100000001000000010000000100000001100110010111000
110011001101000000100000001000000010000000100001100010011001000
1100000010000000100000001000000010000000110101001110000010000000
1000000011010100100000001000000010000000110001001110010011100000
1100010010110100110000001110000000100000001100110011001100000101010000
110001001101000011101000110100011001000111010001100000011010100
10111000110111001110000011100000101101000100000001000000010000000
1100110011100000111001001101000010111000110110001101000010000000
100000001000000011001100110100001011000110110001101000010000000
11100000100000001000000010000000100000001000000010000000101101001
11000001100000010

ARCHIVAGE
NUMÉRIQUE

Un « plan de conservation » numérique ?

- Les informaticiens fournissent un service technique, mais pas la définition du service (définition du besoin)
 - ce n'est pas la même chose de mettre en place un bon système de climatisation, et de savoir quelle est la température et l'hygrométrie idéales d'un magasin.
- Les méthodologies de conservation sont les mêmes pour le papier et le numérique :
 - stratégies préventives,
 - plans de conservation,
 - choix de priorités en fonction de la valeur/fragilité des collections,
 - gestion de risques,
 - alertes et résolutions de problèmes,
 - plans d'urgence...
- Ce sont les bibliothécaires qui doivent établir des niveaux de service



<http://www.flickr.com/photos/nate/284184160/>



<http://www.flickr.com/photos/jurvetson/855530042>

Un « restaurateur » numérique ?

- La conservation du papier, c'est technique
 - (désacidification de masse, encres ferrogalliques, qualité des cuirs, effets des gommes à effacer sur la cellulose du papier...)
- Conserver des documents numériques c'est surtout les préserver
 - (en général, avec le numérique, après = trop tard)
- Stocker/sauvegarder ce n'est pas archiver



Un « magasin » numérique ?

- Le numérique est lui aussi stocké sur un support physique
- L'infrastructure de stockage coûte cher et prend de la place, il faut la sécuriser et la répliquer
- À partir d'une certaine échelle, il faut une couche logicielle pour gérer les fichiers et les données

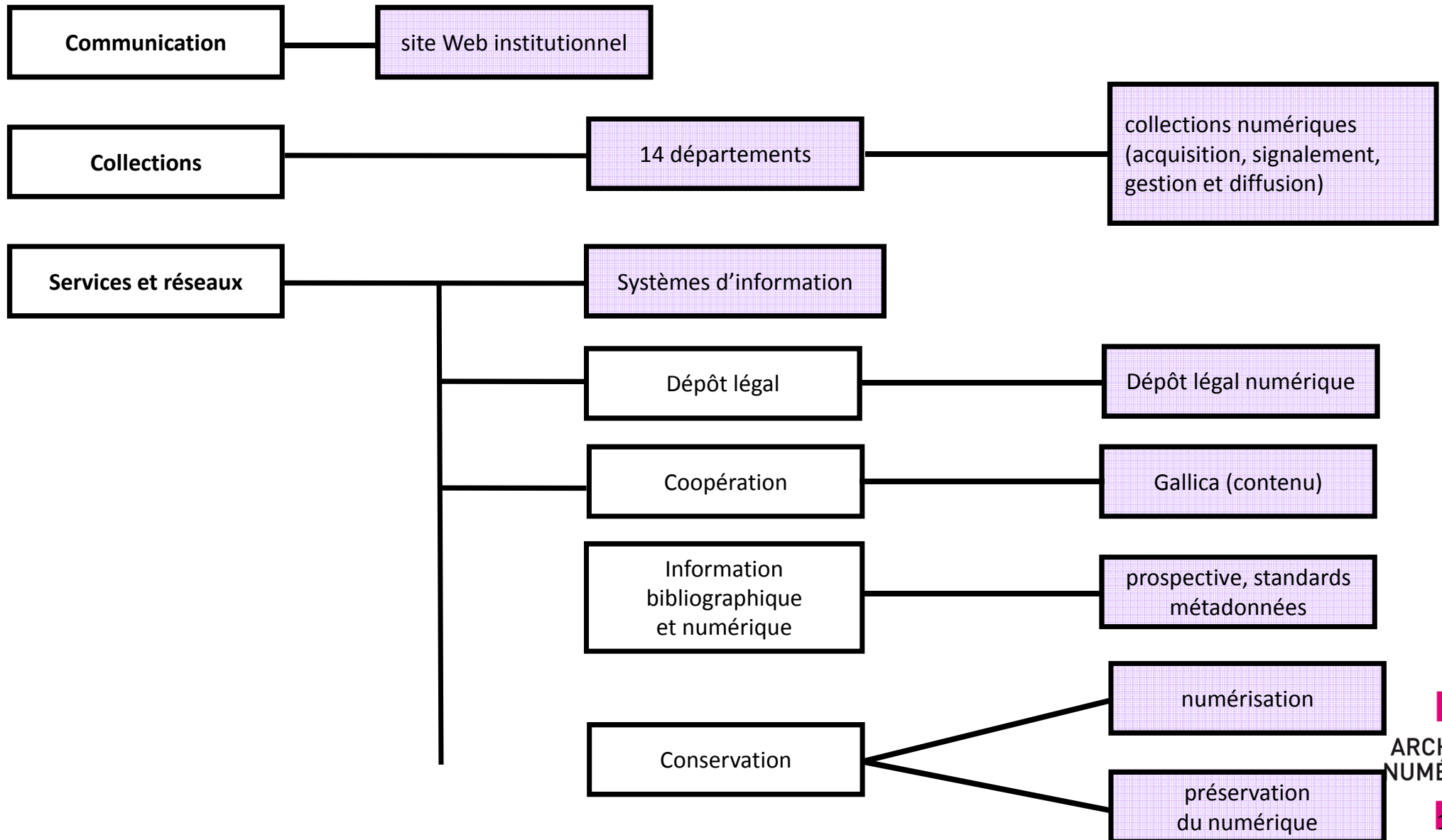


Un « catalogue » numérique ?

- Le catalogage ne se limite plus au descriptif :
 - Métadonnées techniques, administratives (dont droits, historique), etc.
- Heureusement, les parties techniques sont les plus faciles à automatiser : extraction
- On jongle avec plusieurs granularités :
 - d'une logique de document à une logique d'objet

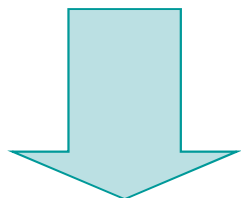
```
- <metadata>
- <oai_dc:dc>
  <dc:identifiant>http://gallica.bnf.fr/ark:/12148/bpt6k10
- <dc:title>
  Trente jours de voyage en zigzag : de Paris à Venise e
  </dc:title>
  <dc:creator>Abadie, Louis (18..-19.. ; avocat)</dc:cre
  <dc:publisher>impr. de A. Lebon (Paris)</dc:publisher>
  <dc:date>1862</dc:date>
  <dc:subject xml:lang="fre">France -- Descriptions et
  <dc:subject xml:lang="fre">Italie -- Descriptions et v
  <dc:language>fre</dc:language>
  <dc:type xml:lang="eng">text</dc:type>
  <dc:type xml:lang="fre">monographie imprimée</dc:typ
  <dc:type xml:lang="eng">printed monograph</dc:typ
  <dc:rights xml:lang="fre">domaine public</dc:right
  <dc:rights xml:lang="eng">public domain</dc:rights
  <dc:format>application/pdf</dc:format>
  </oai_dc:dc>
</metadata>
</record>
- <record>
- <header>
  <identifiant>oai:bnf.fr:gallica/ark:/12148/bpt6k39258m</
  <datestamp>2007-01-19</datestamp>
  <setSpec>gallica</setSpec>
  <setSpec>gallica:9</setSpec>
  <setSpec>gallica:9:944</setSpec>
  <setSpec>gallica:monographies</setSpec>
</header>
- <metadata>
- <oai_dc:dc>
  <dc:identifiant>http://gallica.bnf.fr/ark:/12148/bpt6k39
  <dc:title>
  Bâtiments de la Bibliothèque nationale : rapport / M.
  </dc:title>
  <dc:contributor>
  Barthélemy-Saint-Hilaire, Jules (1805-1895). Rédact
  </dc:contributor>
  <dc:creator>France. Ministère de l'instruction publicu
  <dc:publisher>Impr. nationale (Paris)</dc:publisher>
  <dc:date>1879</dc:date>
  <dc:subject xml:lang="fre">Bibliothèque nationale (F
  <dc:language>fre</dc:language>
```

Résultat, du numérique partout



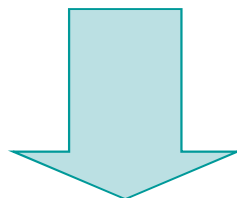
Quel impact sur les métiers ?

Gestion des collections numériques



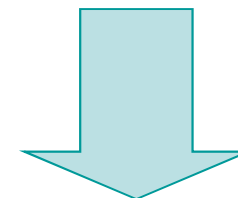
Expertise des chargés
de collection +
producteurs
d'information
numérique

Expertise de préservation



Expertise de
conservation +
cohérence des
données numériques

Gestion du magasin numérique



Expertise informatique
+ nouvelle fonction de
magasinier numérique
?

SPAR est un système

- conçu selon des **normes**
 - norme OAIS (ISO-14721)
 - Modèle de référence pour un Système ouvert d'archivage d'information
 - norme NF Z42-013 en certains de ses principes
 - Spécifications relatives à la conception et à l'exploitation de systèmes informatiques en vue d'assurer la conservation et l'intégrité des documents stockés dans ces systèmes
 - spécifications RDF du W3C pour l'interrogation des données
- de **confiance**
 - répondant aux exigences les plus hautes : agrément **SIAF** (mars 2013)
 - prêt pour la certification ISO 16363:2012
 - en conformité avec les recommandations de la CNIL relatives à la protection des données personnelles (audit automne 2011)



SPAR : un outil au service de la préservation

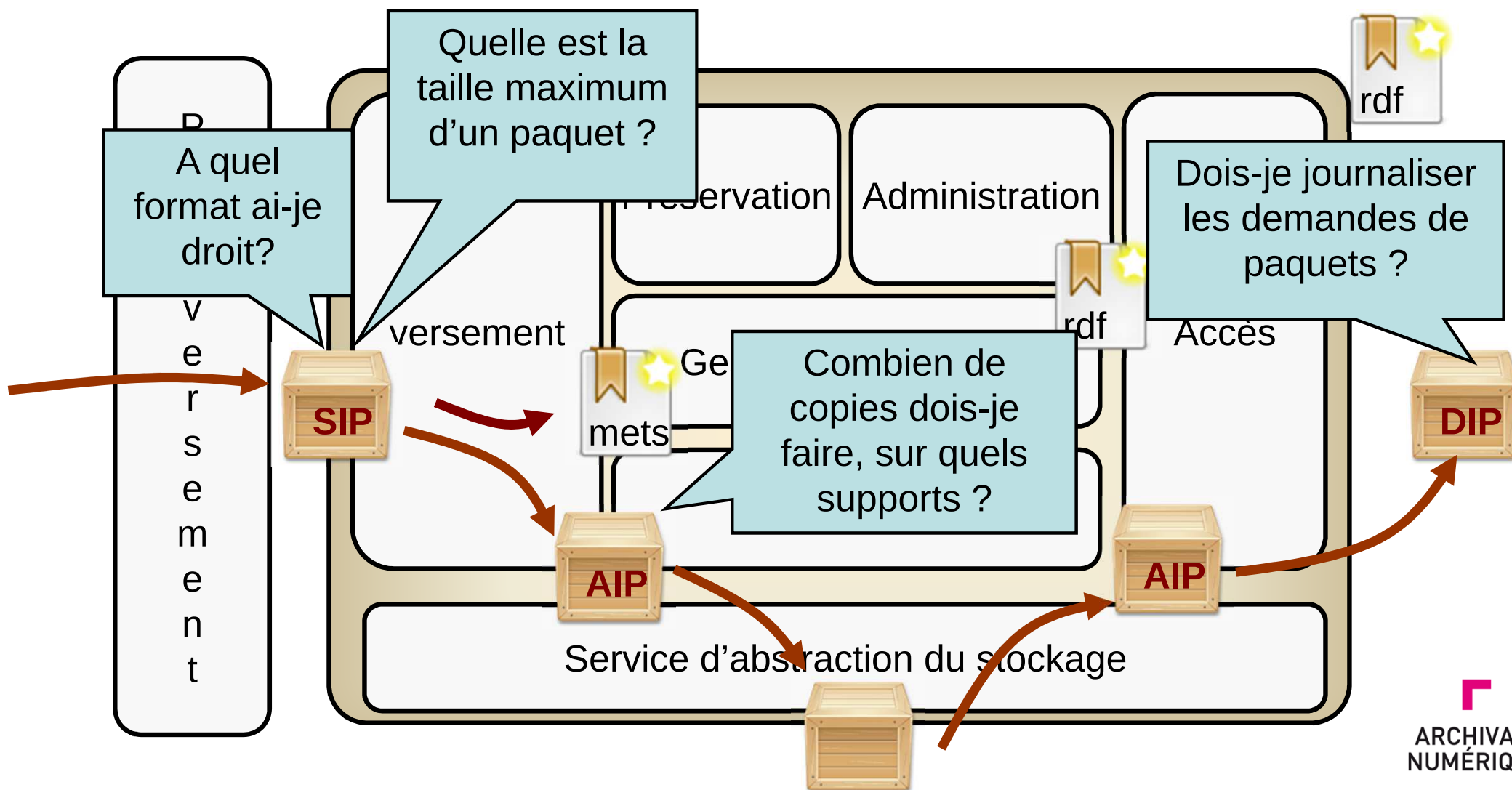
- Données redondées, sur 2 sites situés en France, dans des zones en accès restreint, avec cloisonnement des serveurs, plan de reprise d'activité
- Contrôle à l'entrée les données reçues (AQS)
- Assure une surveillance continue des données et des équipements
- Gère les habilitations
- Fournit un horodatage de toute action
- Permet la veille automatisée sur les formats de données versées.
- Permet les transformations de formats : « opération de migration » pour garantir ainsi l'accès à long terme des fichiers.
- Assure la restitution des données
- Experts assurent le gestion et le développement du système



Les accords de qualité de service

- Les accords de qualité de service : contracter avec les utilisateurs pour un système plus transparent
- Le système n'est pas une boîte noire maîtrisée uniquement par les experts techniques.

AQS : prise en compte dans le système



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

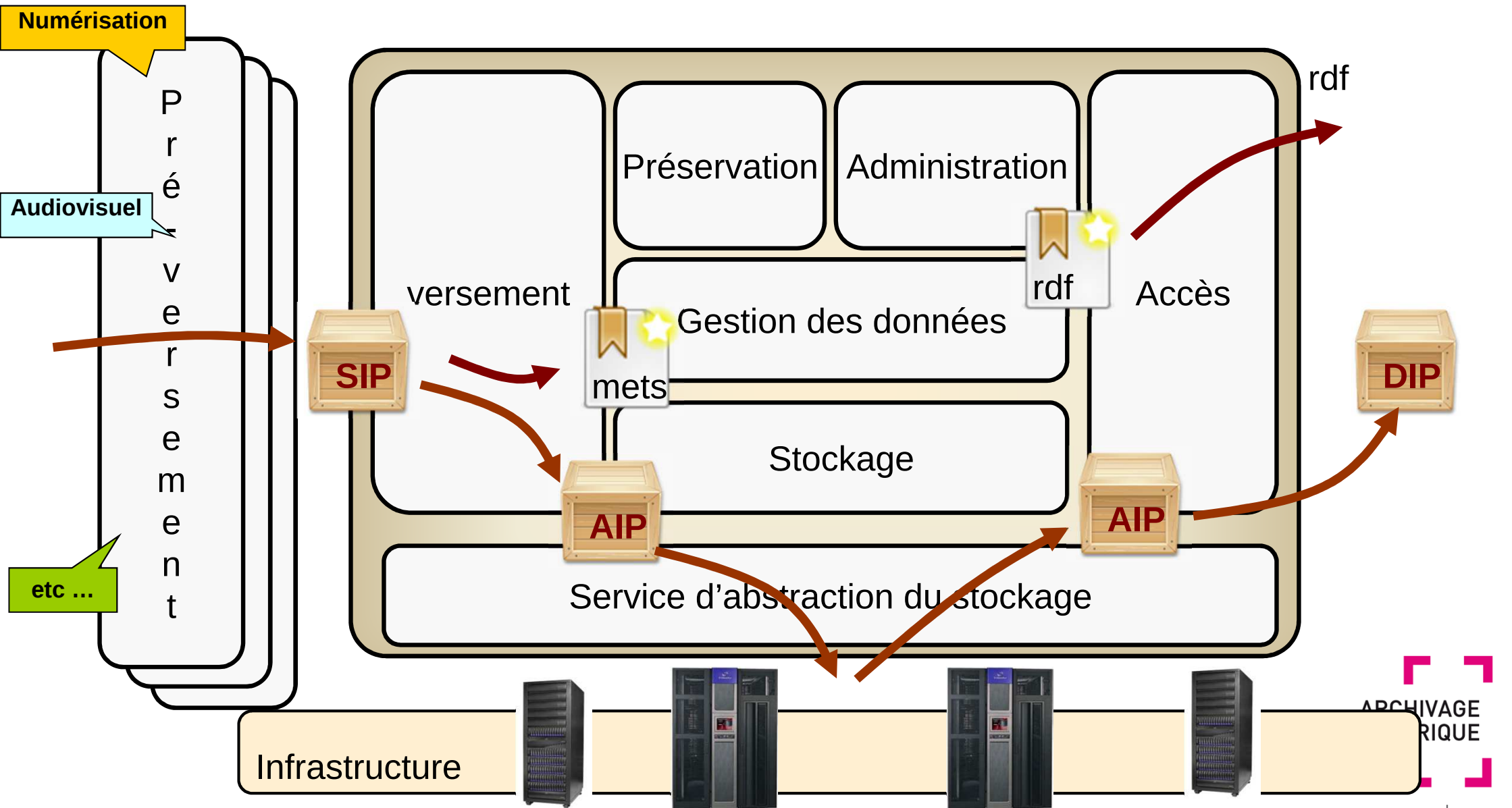
Contrat de service.pdf

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

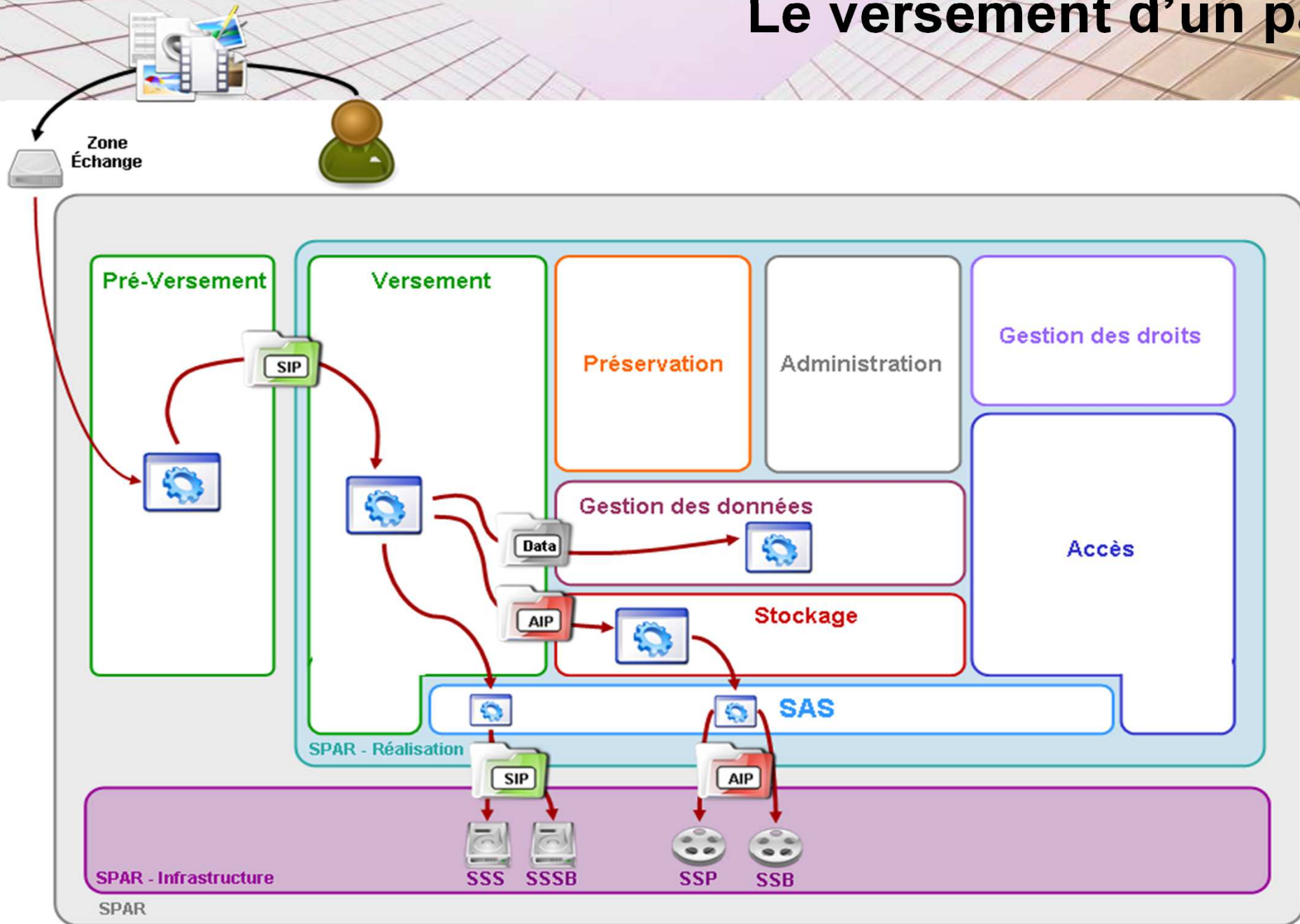
La filière et leurs chaînes

- 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

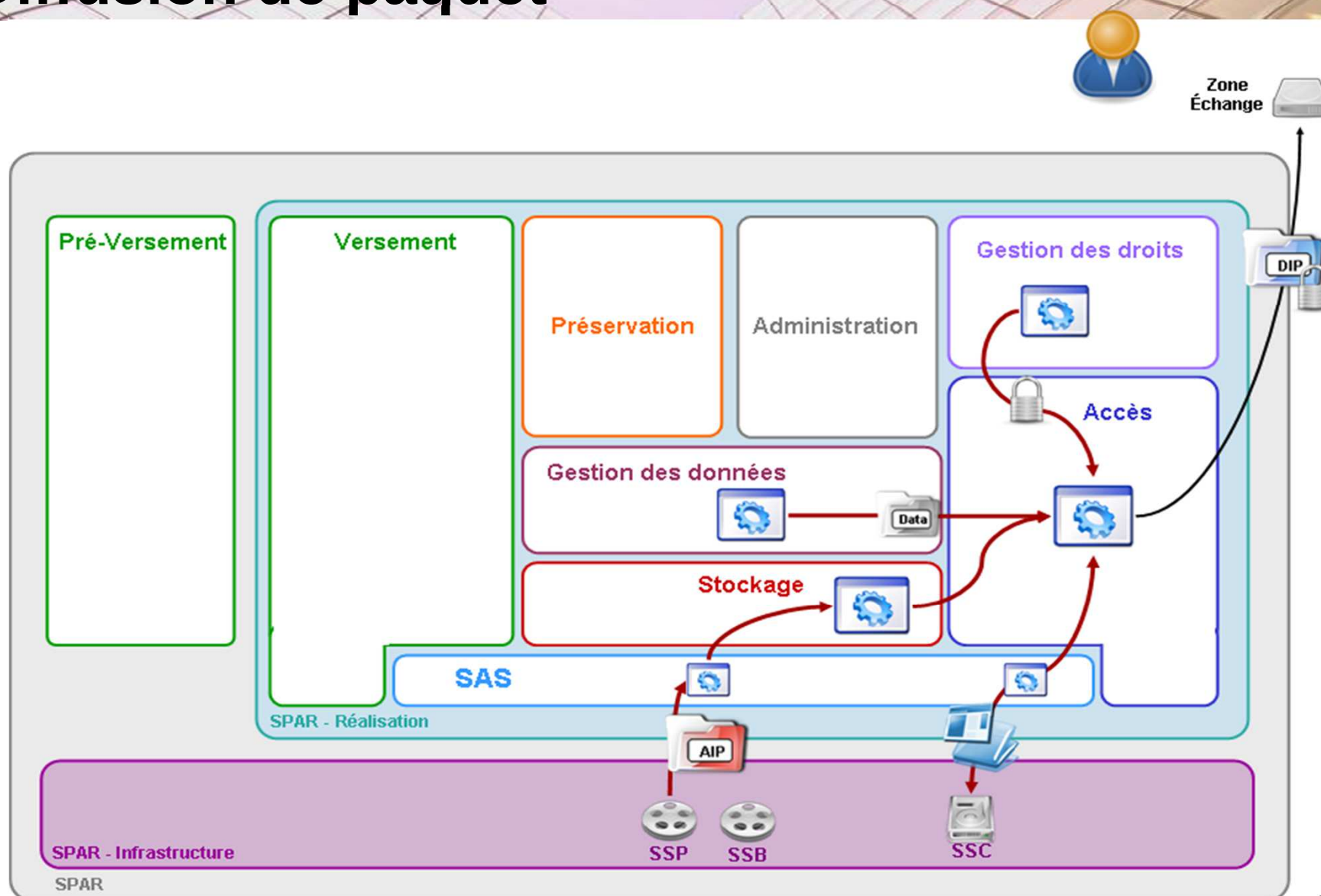
Les modules de SPAR



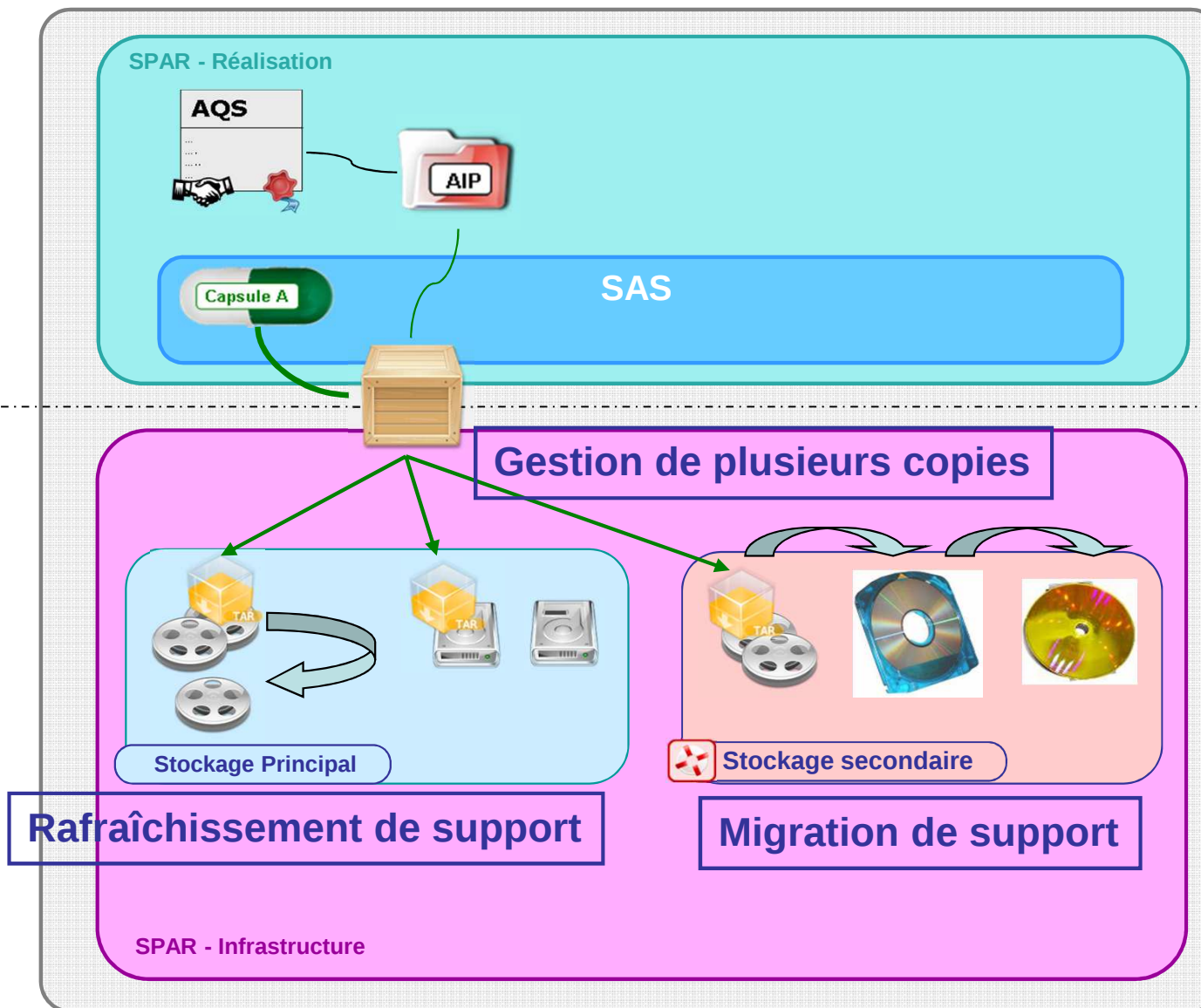
Le versement d'un paquet



Diffusion de paquet



L'abstraction de stockage



Vue Applicative

Vue Infrastructure

Risques

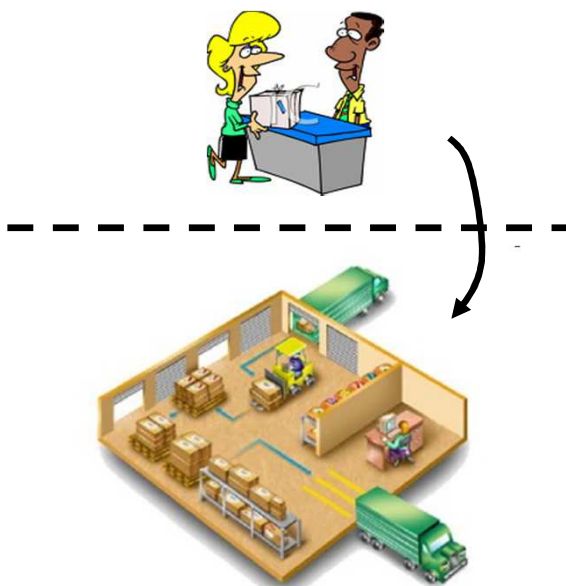


Enjeux

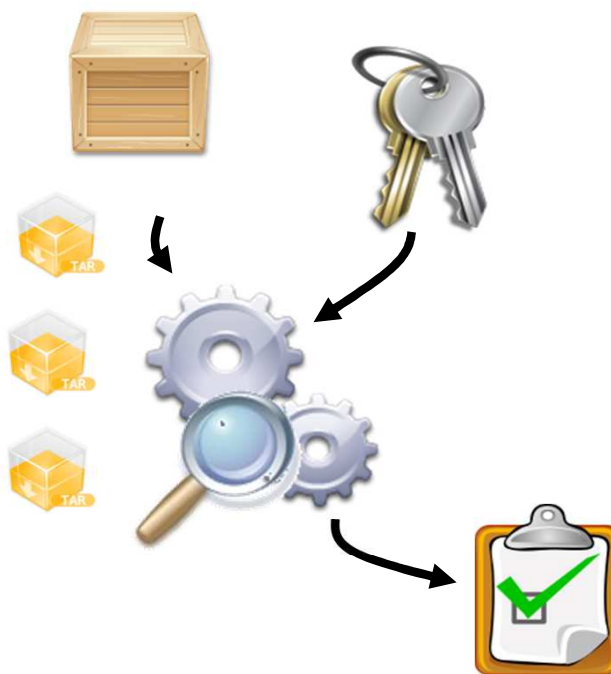
- Garantir l'intégrité des enregistrements
- Garantir la pérennité du stockage
- Gérer l'obsolescence technologique
- Automatiser les différentes répliques et copies des enregistrements
- Assurer les audits des enregistrements
- Assurer les changements de l'infrastructure de manière transparente
- Garantir une indépendance vis-à-vis des constructeurs
- Dissocier stockage logique et physique
- Gérer les sites distants
- ...

Rôles et responsabilités

Abstraction du stockage



Audit des enregistrements



Rappel OAIS

- Entité Stockage [4-7]
- Remplacer les supports
 - Contrôle d'erreurs
- Migration numérique [5-2 à 10]
- Rafraîchissement de support
 - Duplication
 - Ré-empaquetage

Migration transparente de support



Un système ouvert permettant un service pour des tiers

- Dès sa conception, le système SPAR a été conçu pour accueillir un **service d'archivage** pour des données extérieures.



- Des collections numérisées de nombreux partenaires de Gallica déjà archivées
- Plusieurs demandes d'archivage numérique pour des tiers :
 - un contrat signé : centre Pompidou virtuel
 - d'autres contrats en cours.
- L'ouverture permet une mutualisation des coûts.

La fixation des tarifs

- 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).
- Le coût est confirmé à partir de l'accord de qualité de service.



Prix de revient lissé sur 5 ans

BnF - Archivage numérique Tarif général annuel, au téra octet, en € HT, basé sur 5 ans		
Volume des données à archiver sur 2 bandes :	extranet	sur-mesure
pour 1 To	2 545 €	3 990 €
de 2 à 5 To	1 960 €	2 680 €
de 6 à 9 To	1 740 €	2 160 €
de 10 à 29 To	1 590 €	1 840 €
de 30 à 49 To	1 420 €	1 530 €

Options

Archivage complémentaire sur 1 disque : + 315€

Archivage complémentaire sur 2 disques : +630€

La suite...

- Tarifs validés en conseil d'administration en fin d'année 2012.
- Prévoient réductions pour les partenaires Gallica
- Service proposé depuis le 1^{er} trimestre 2013 et décrit en ligne (présentation, contrat, tarifs) :
- http://www.bnf.fr/fr/professionnels/numerisation/a.bnf_archivage_numerique.html

Pour les professionnels

Actualités
professionnelles

Dépôt légal

ISSN, ISBN, autres
numéros

Site de la
Bibliographie
nationale française

Catalogage et
indexation

Récupération des
données

Coopération
nationale

Formations et
journées
professionnelles

Action
internationale

Numérisation et
archivage
numérique

Sélectionner :
politique et
programmes

Gérer les droits

Numériser

Conserver : SPAR

Mettre en ligne :
l'accès aux
documents

Conserver pour des
tiers : BnF - Archivage

Conserver pour des tiers : BnF - Archivage numérique

La BnF ouvre son système d'archivage numérique SPAR à d'autres organisations ; celui-ci permet la préservation sécurisée de leurs données à très long terme.

- ✦ Les enjeux de la mutualisation de l'archivage
- ✦ Le service BnF - Archivage numérique

Les enjeux de la mutualisation de l'archivage

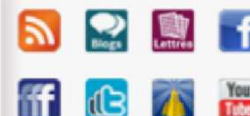


Salle des serveurs sécurisés

La conservation des données numériques par la BnF au titre de ses missions fondamentales pose de façon cruciale la question de la **pérennisation**. Sommes-nous certains que, dans dix ans, une période somme toute assez courte, les documents produits seront encore consultables et que les efforts consentis à leur création ne seront pas définitivement perdus ? Certaines organisations font déjà état d'une perte de leurs données.

L'absence de stratégie dans ce domaine conduit inéluctablement au pire. Si les coûts de stockage informatique ne cessent de décroître, cela ne doit pas faire illusion. L'archivage à long terme ne se résume pas au stockage mais nécessite la mise en place d'un dispositif plus complexe.

a+ | a-



Écouter la page :



Contact

Bibliothèque nationale
de France
Direction des Services
et réseaux
courriel : [archivage-
numerique@bnf.fr](mailto:archivage-numerique@bnf.fr)

En savoir plus

- ✦ **Présentation de BnF - Archivage numérique** [fichier .pdf – 1369 Ko – 30/01/13 – 33 p.] ppt converti en pdf
- ✦ **Conserver : SPAR**

Consulter

- ✦ **Contrat type** [fichier .pdf – 233 Ko – 30/01/13 – 15 p.]
- ✦ **Nos tarifs** [fichier .pdf – 33 Ko – 10/04/13 – 2 p.]

ARCHIVAGE
NUMÉRIQUE



Question ?