

Pérenniser 2 Po de documents numériques

**Systeme
de Préservation et d'Archivage Réparti
(SPAR)**

14 avril 2016

{ BnF

Ordre du jour

- BnF, bref historique et missions
- Préservation numérique à la BnF
- SPAR, le projet
- Modèles de données et flux principaux
- Gérer la masse
- Organisation autour de SPAR

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

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...



Ordre du jour

- BnF, bref historique et missions
- Préservation numérique à la BnF
- SPAR, le projet
- Modèles de données et flux principaux
- Gérer la masse
- Organisation autour de SPAR

SPAR en chiffres (au 05/04/2016)

- Ouverture de l'exploitation : mai 2010
 - Préservation de la numérisation courante de la BnF
 - Stockage répliqué sur deux sites
- 5 512 000 paquets archivés
- 233 509 000 objet-données (fichiers)
- > 2,35 Po (= 2 350 To) de données brutes
- > 3,1 milliard de triples (métadonnées élémentaires)
- Des manuscrits très volumineux ont été versés (paquets de taille supérieure à 100 Go)
- Première migration technologique réalisée : nouvelle génération de bande
- Filières disponibles : monographies, périodiques, lots d'image, audio, vidéo, archivage du web, documents bureautiques, objets 3D, tiers archivage

Le projet SPAR : contexte

Applications de production de données



Numérisation de conservation



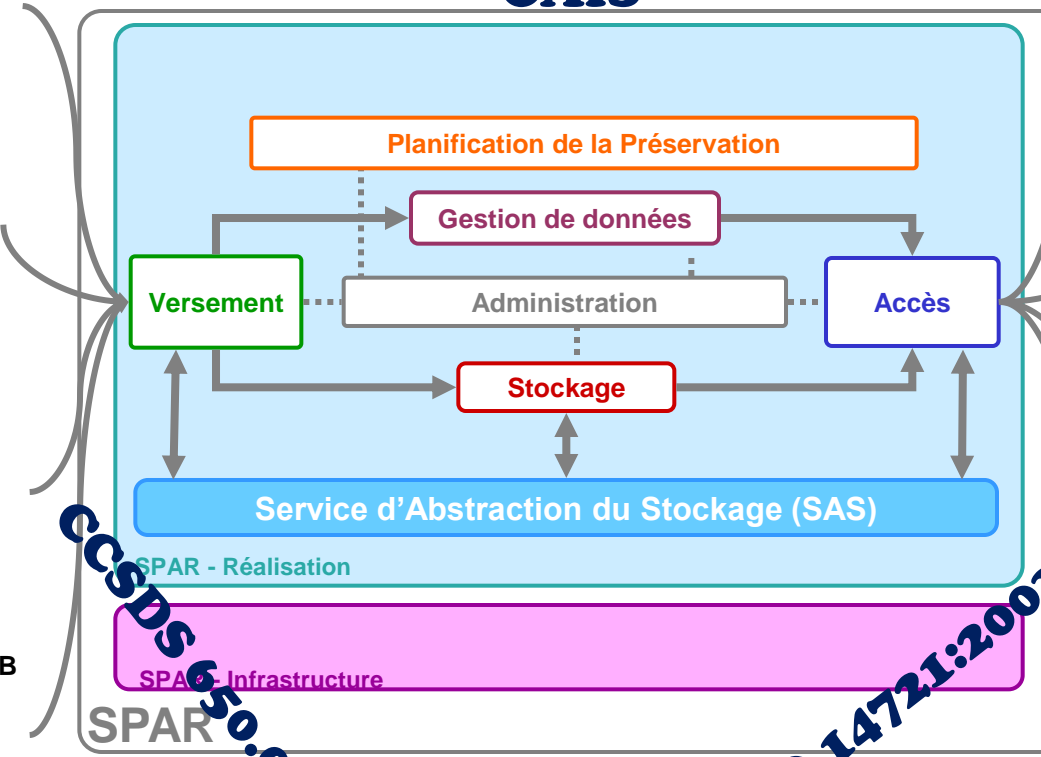
Production administrative
Record Management



Archives du WEB

...

OAIS

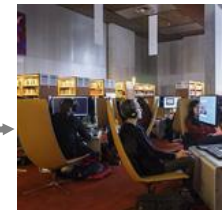


CCSDS 650.0-M-2

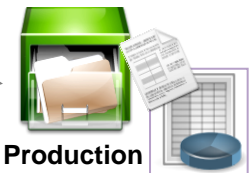
ISO 14721:2003



Gallica



Postes publics



Production administrative



wayback

Applications de diffusion de données

SPAR

Enjeu de la préservation

Voici un document numérique que j'ai créé en 1998.

De quoi s'agit-il, déjà ?
Est-ce bien ce qui est indiqué sur la disquette ?

Métadonnées descriptives

Je l'ai enregistré sur une disquette, est-elle toujours en bon état ?

Support

Ça marche !

Mais j'ai perdu toute ma mise en forme ...

Authenticité / Intégrité



Mon portable, acheté en 2004, n'a pas de lecteur de disquette. Où en trouver un ?

Environnement matériel

J'ai trouvé le logiciel, mais puis-je l'installer et l'utiliser sous Windows XP ?

Système d'exploitation

J'ai créé ce document avec Claris Works. Comment vais-je retrouver ce logiciel ?

Environnement logiciel /
Formats du fichier

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^{er} août 2006 ou du dépôt légal des documents audiovisuels). 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

Numérisation des plaques de verre du fonds Rol



Source gallica.bnf.fr / Bibliothèque nationale de France

Voir http://www.bnf.fr/fr/professionnels/anx_actu_conservation/a.cn_act_num29_art2.html

Enjeu technique : l'explosion du numérique

- De la numérisation de valorisation à la numérisation de préservation

- Création d'une bibliothèque numérique qui monte, qui monte...
- Disparition progressive des moyens de productions de microformes
- Numérisation suite à l'obsolescence voire la disparition des équipements de restitution



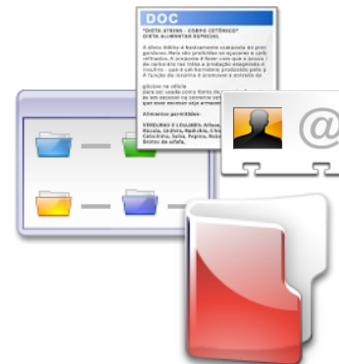
- Dépôt légal de substitution

- Affiches de grand format, presse quotidienne régionale



- Documents nés numériques

- Dépôt légal du Web (1^{er} août 2006)
- Production administrative
- E-books



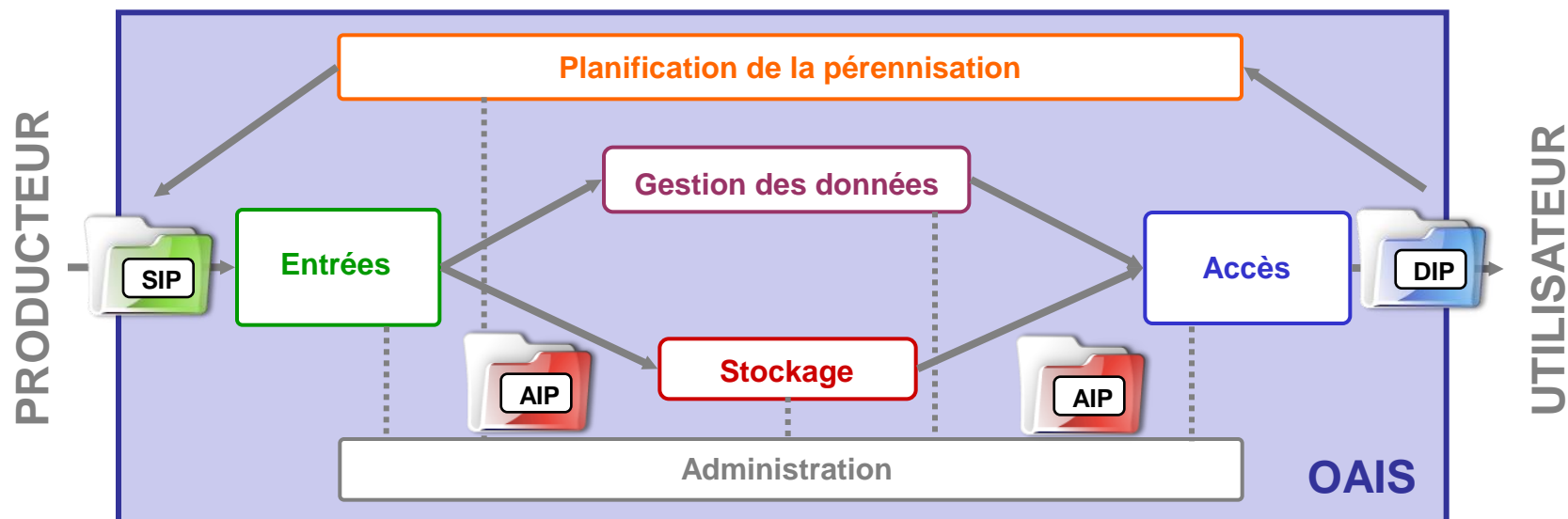
Avec une masse à gérer

- qui croit rapidement
- dont le stockage peut être dispersé
- qui est soumise aux évolutions techniques rapides
- et aux parcs et solutions informatiques hétérogènes

Environnement institutionnel et normatif

- BnF, établissement public administratif dépendant du Ministère de la Culture et de la Communication
- Dès 2004, la BnF a entamé la réflexion sur la mise en œuvre d'un système de préservation du numérique à long terme.
- Elle s'appuie sur les normes internationales et plus particulièrement sur la norme OAIS (ISO-14721:2003)
 - Modèle de référence pour un Système ouvert d'archivage d'information
- Prise en compte de certaines directives de la norme NF Z42-013
 - 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
- La BnF est pour ces propres archives, y compris numériques, sous la tutelle scientifique et technique du SIAF (code du patrimoine)

Modèle OAIS (norme ISO 14721:2003)



- À quoi sert-il réellement ?
 - À établir la **responsabilité de la conservation** : notion de contrat
 - À **surveiller les documents** avant que les dégradations ne se produisent (en général, avec le numérique, «après» = trop tard)
 - À **garantir l'intégrité des documents**, vérifier qu'ils ne sont pas altérés, et s'ils le sont, expliquer pourquoi
 - À **organiser l'activité de conservation** autour d'un référentiel commun pour faciliter les évolutions

Ordre du jour

- BnF, bref historique et missions
- Préservation numérique à la BnF
- SPAR, le projet
- Modèles de données et flux principaux
- Gérer la masse
- Organisation autour de SPAR

Objectif SPAR

- SPAR : Système de Préservation et d'Archivage Réparti
- Un système conçu selon la **norme OAIS**
- Un système qui **garantit**
 - la pérennisation des données numériques
 - et l'accès à ces données
- Un système centralisé qui prend en compte la **diversité** des données numériques archivées (réparties en **filières**)
- Un système de confiance **agréé par le SIAF**
- Un système capable de fournir un service de **tiers-archivage** à des partenaires

Démarche utilisée pour SPAR

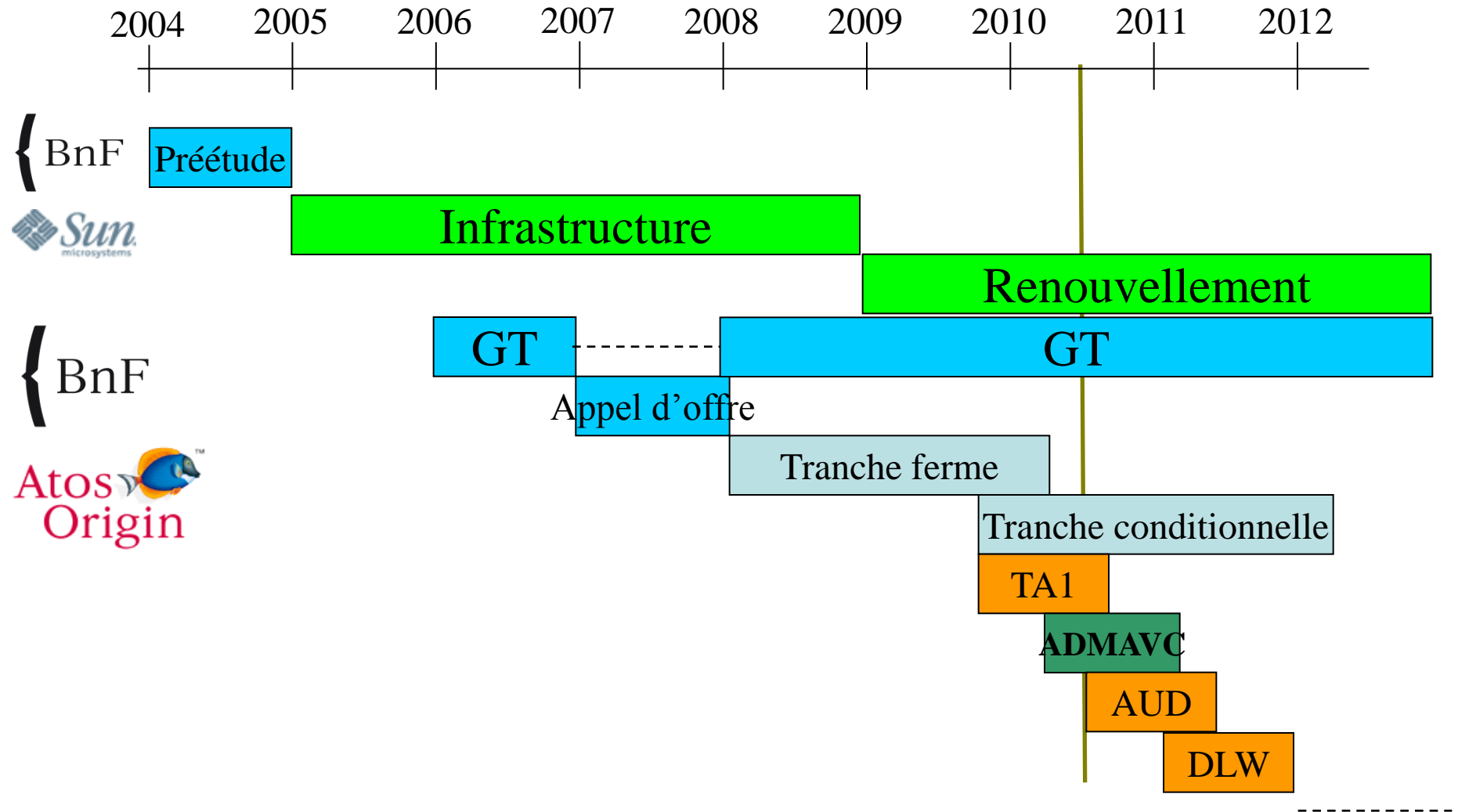
- Définition du besoin (Groupes de travail) 2006
- Étude technique des solutions possibles 2006
- Validation du scénario de réalisation par le comité de pilotage début 2007
- Marché de réalisation 2007
 - Rédaction du cahier des charges de réalisation
 - Réception des offres
- Réalisation par itération 2008 ...

Groupes de travail

- Les groupes de travail fonctionnels : utilisateurs représentatifs des départements de la BnF
 - G1 : groupe transverse de suivi
 - G2 : communauté d'utilisateurs
 - G3 : modèles d'information
 - G4 : gestion des risques
- Groupe de travail technique (DSI)

Planning

mise en production
mai 2010



{ BnF



{ BnF



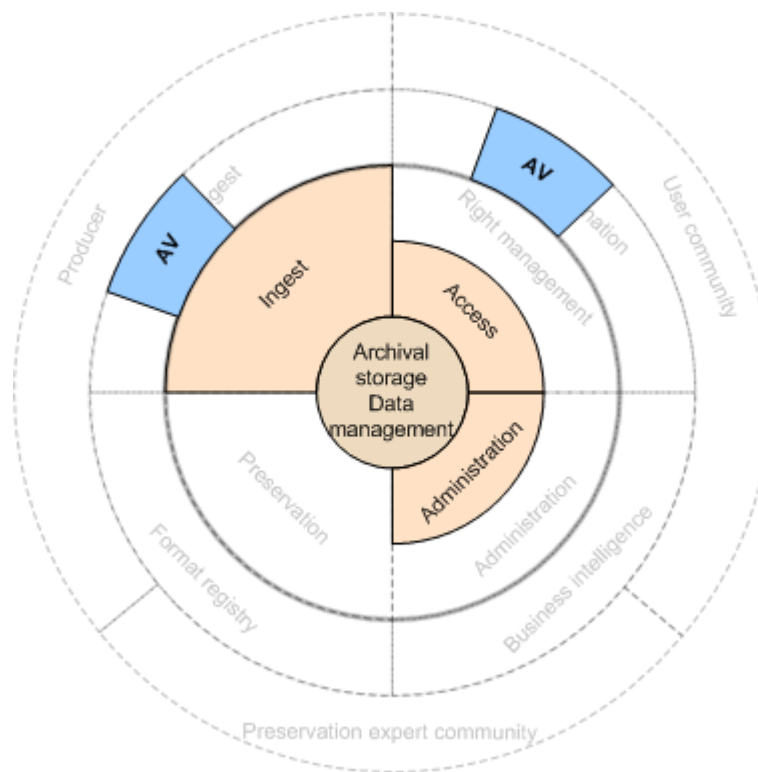
{ BnF

SPAR

Réalisation : démarche itérative



Tranche ferme

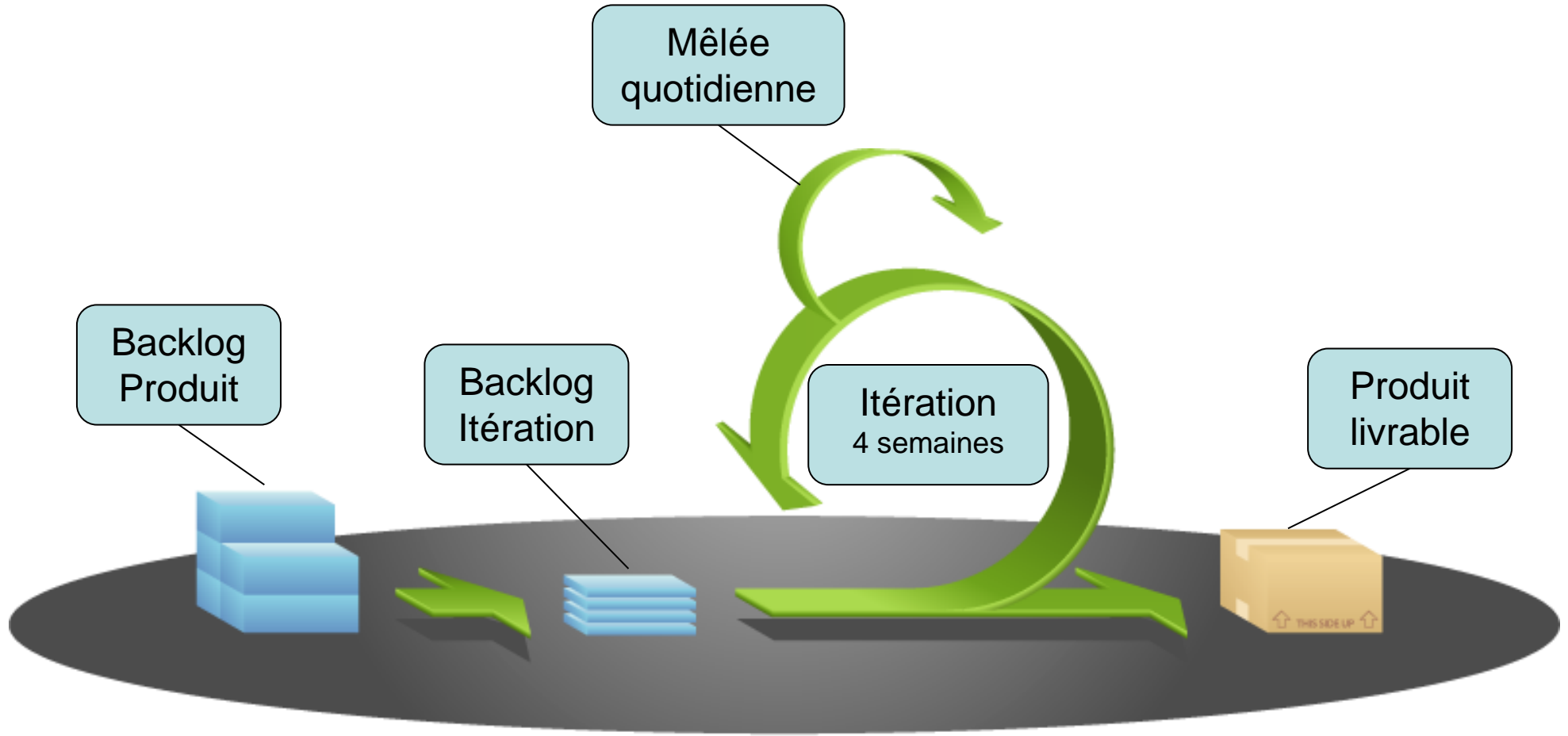


Tranche conditionnelle

Marché d'évolutions

- **Marché initial 2008-2012**
 - Tranche ferme forfaitaire sur 2 ans
 - Tranche conditionnelle à bons de commande sur 2 ans
- **1^{er} marché d'extensions et d'évolutions**
 - 2012-2015
 - Marché à bons de commande
- **2^{ème} marché d'extensions et d'évolutions**
 - 2015-2018
 - Marché à bons de commande

Poursuite de la réalisation : méthode agile

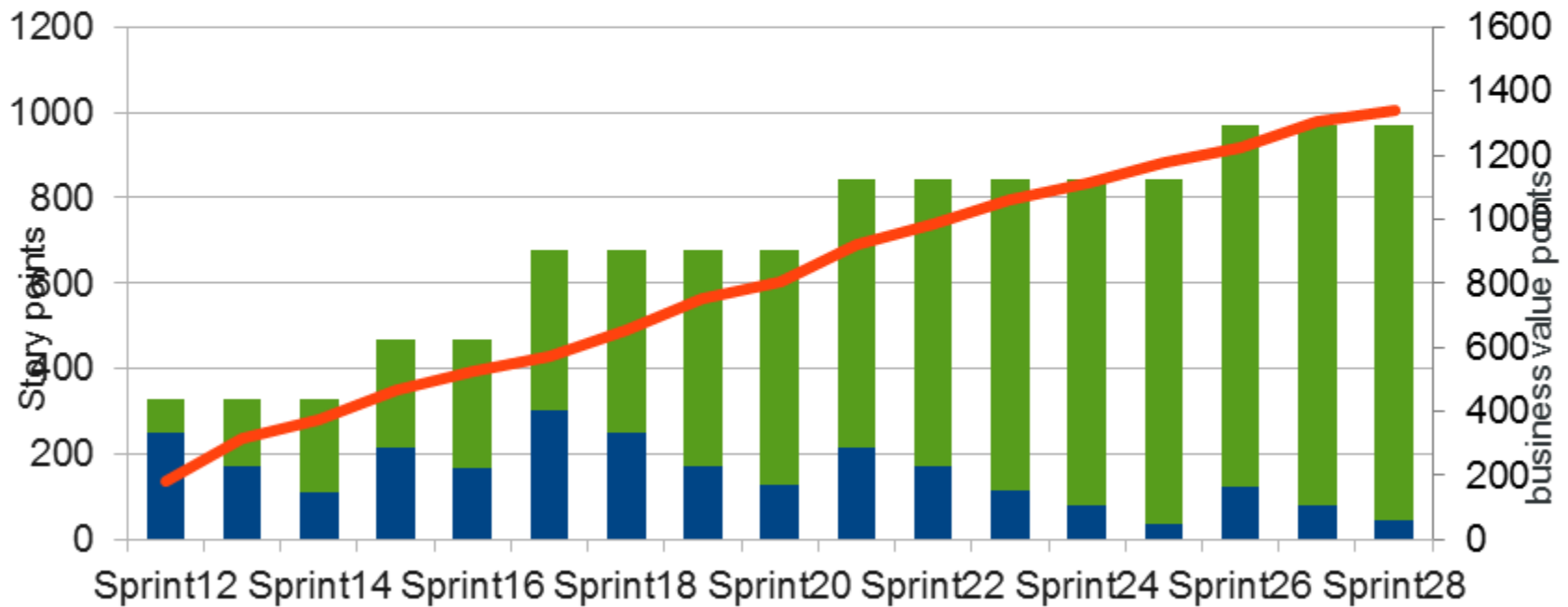


COPYRIGHT © 2005, MOUNTAIN GOAT SOFTWARE

Poursuite de la réalisation : méthode agile

Legend:
■ increased points
■ Remaining points
— increased BV

Burndown points charts & burnup BV chart
BC5, BC6, BC7, BC8, BC9 et BC10



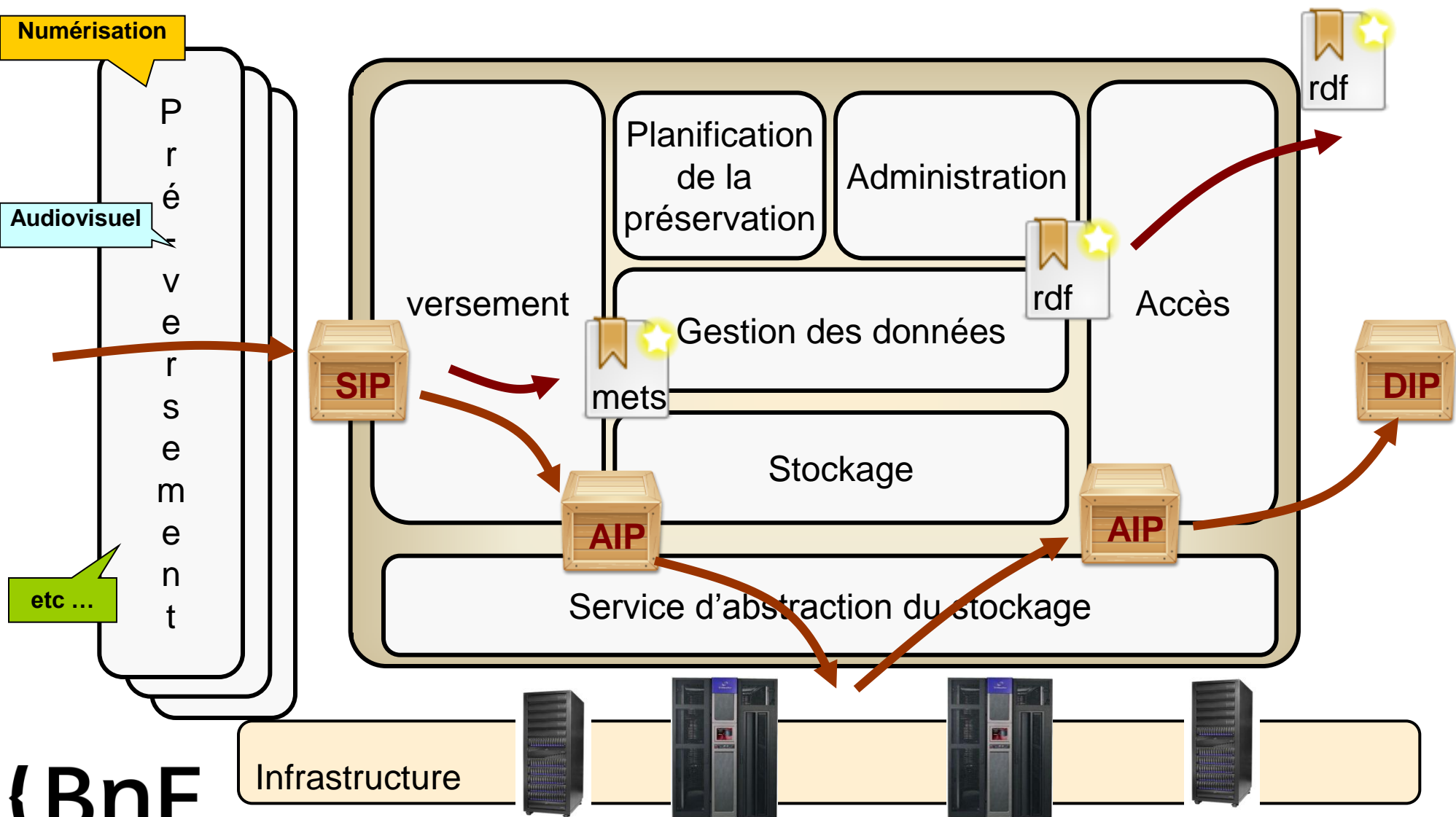
Ordre du jour

- BnF, bref historique et missions
- Préservation numérique à la BnF
- SPAR, le projet
- Modèles de données et flux principaux
- Gérer la masse
- Organisation autour de SPAR

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)


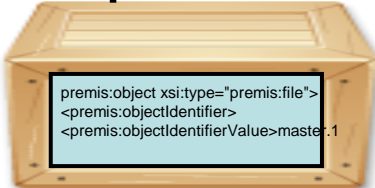
Architecture générale



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**

Que préserve-t-on?

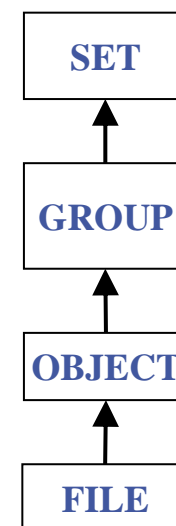
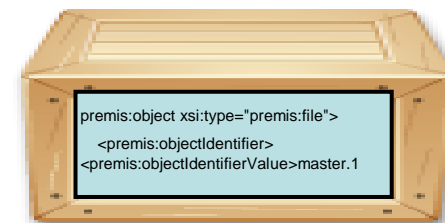
- Un document numérisé ou né numérique c'est :
 - Une description (métadonnées) : titre, auteur, date...
 - Des fichiers texte, image, ... 
 - Une structure physique (les fichiers sont ordonnés)
 - Une structure numérique
 - Les formats des fichiers
 - Les opérations effectuées...
- Tout cela dans un 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



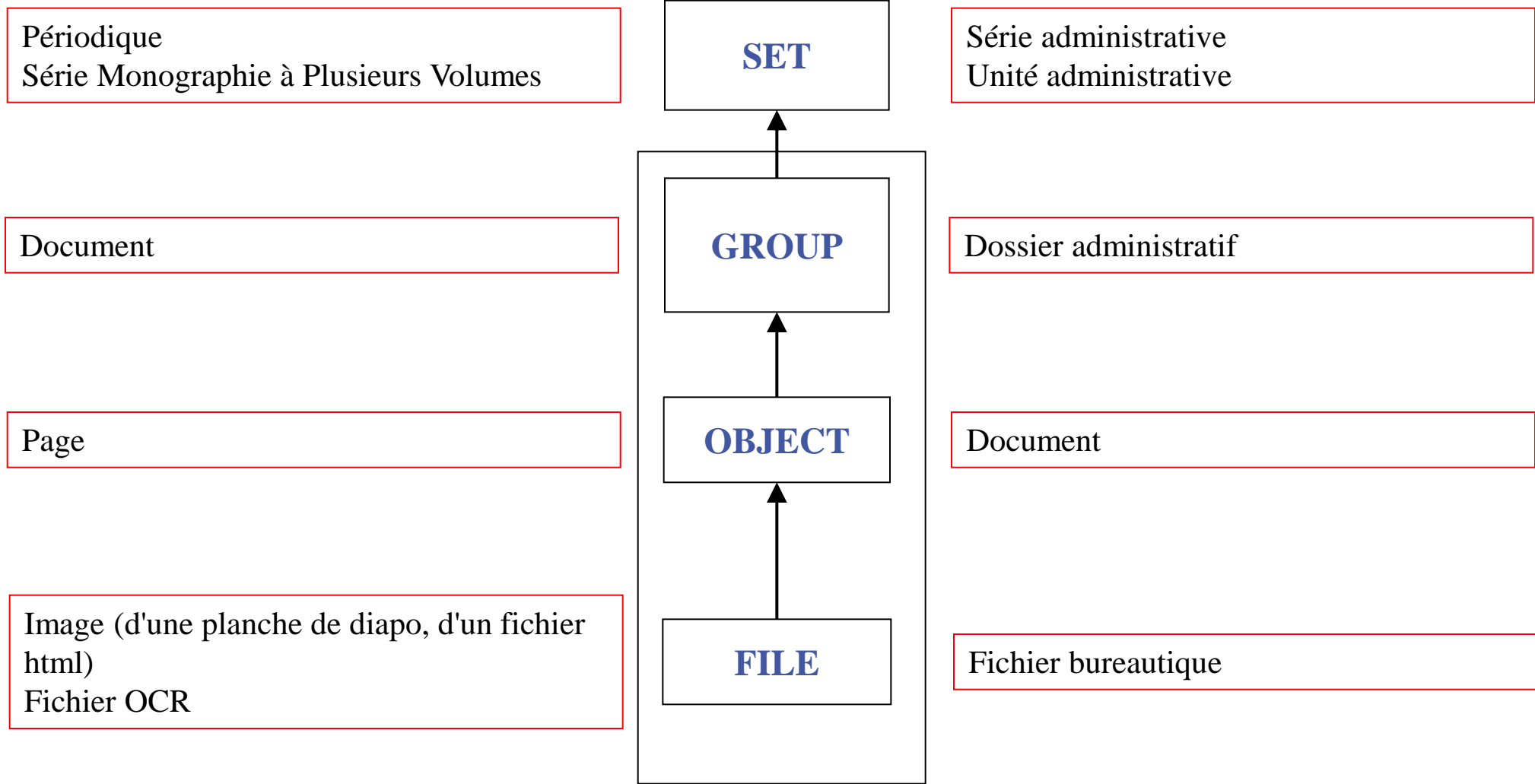
```
premis:object xsi:type="premis:file">  
<premis:objectIdentifier>  
<premis:objectIdentifierValue>master.1
```

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



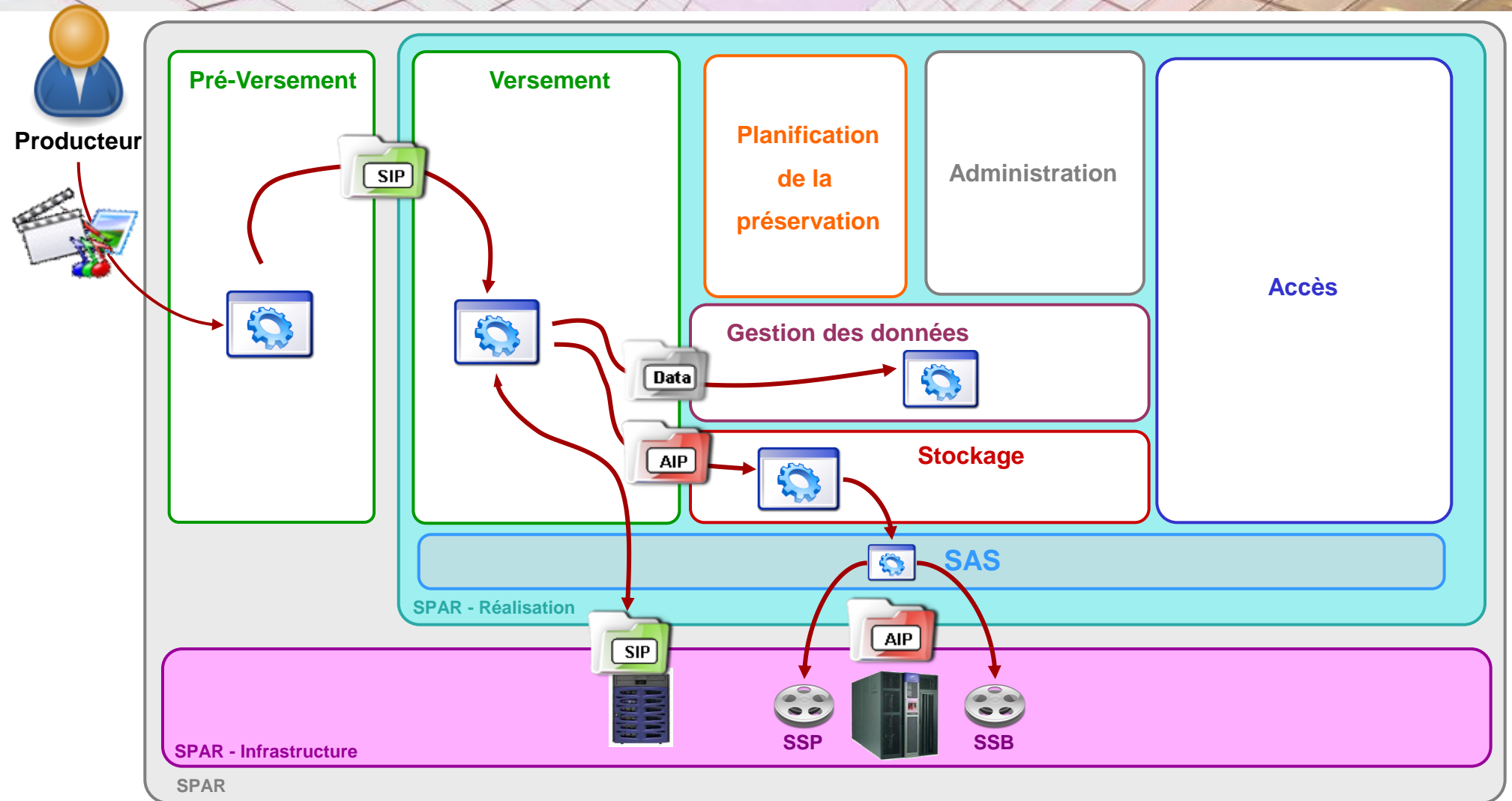
La granularité des objets dans SPAR



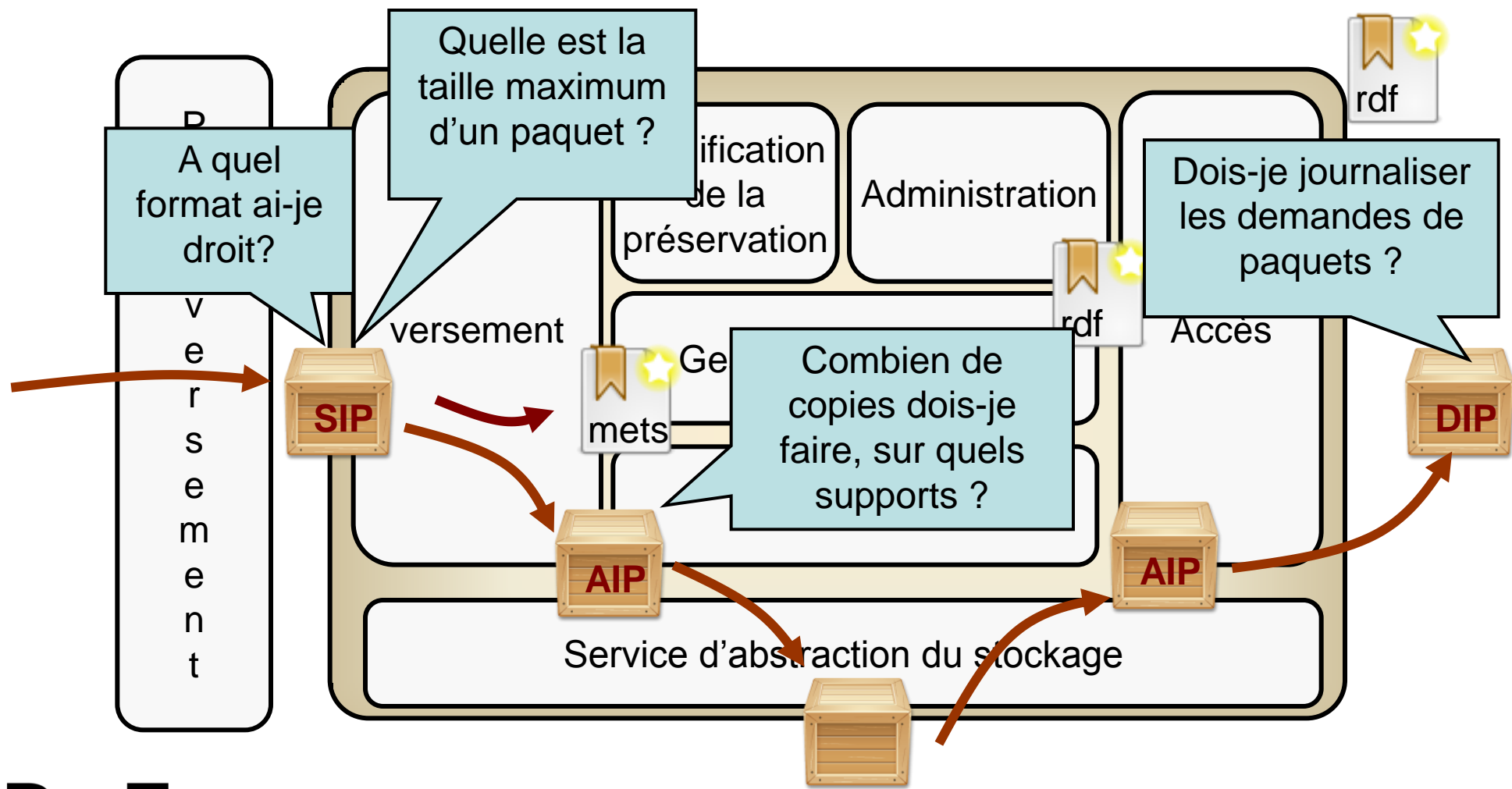
Les schémas de métadonnées utilisés dans SPAR

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...)	

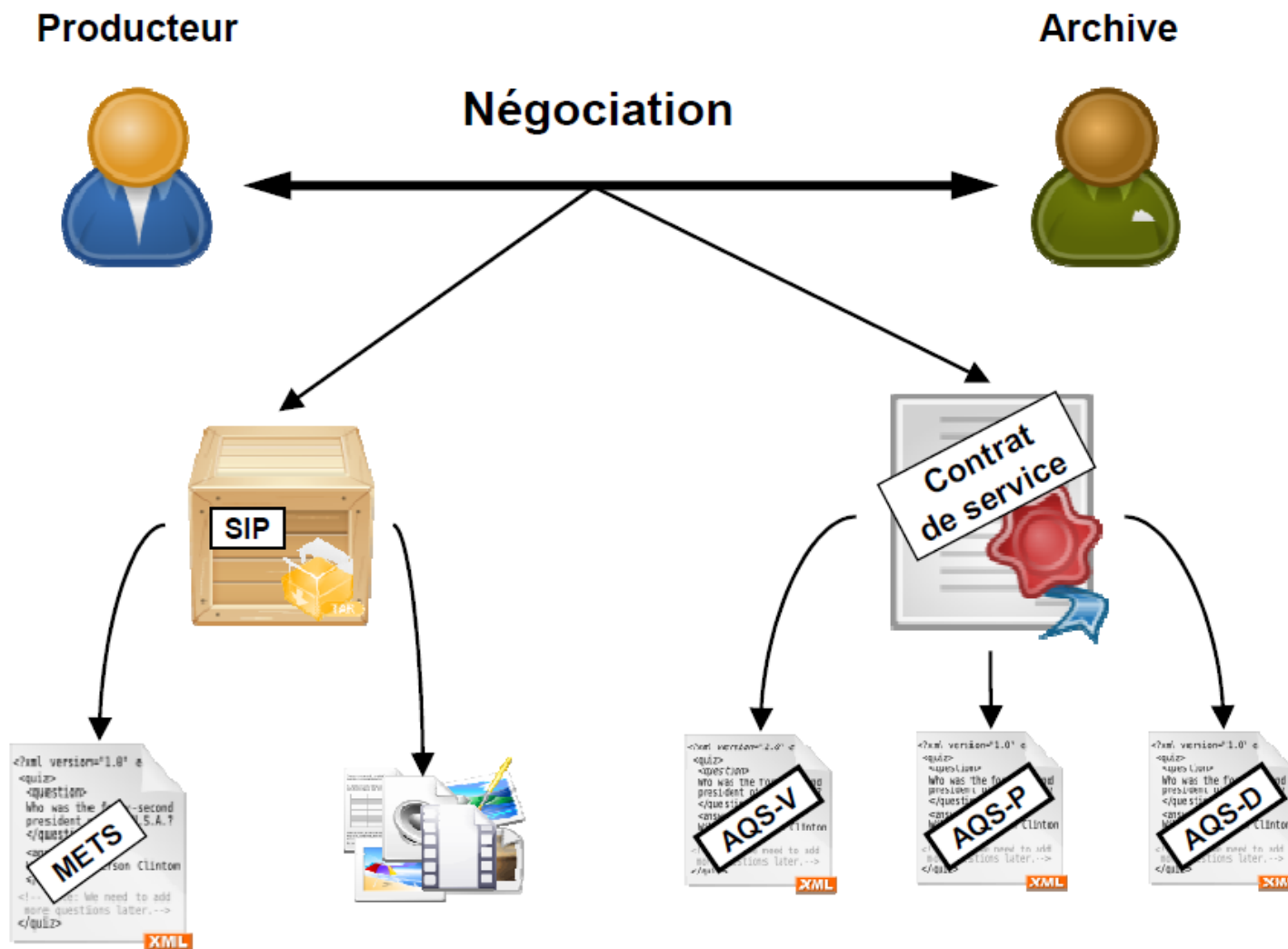
Le versement d'un paquet



Accords de Qualité de service



Principe : contractualiser la filière



Exemple d'AQS pour la chaîne de Numérisation

- N'autoriser que les formats suivants :
 - html, xhtml, tiff (noir/blanc, niveau de gris, 24 bits), jpeg 8 et 24 bits, format xml résultant de la reconnaissance optique et de la table des matières.
- Opérationnel 7j/7
- Taille maximale de 200 Go
- 8 000 fichiers au maximum par paquet
- 10 000 paquets au maximum par jour

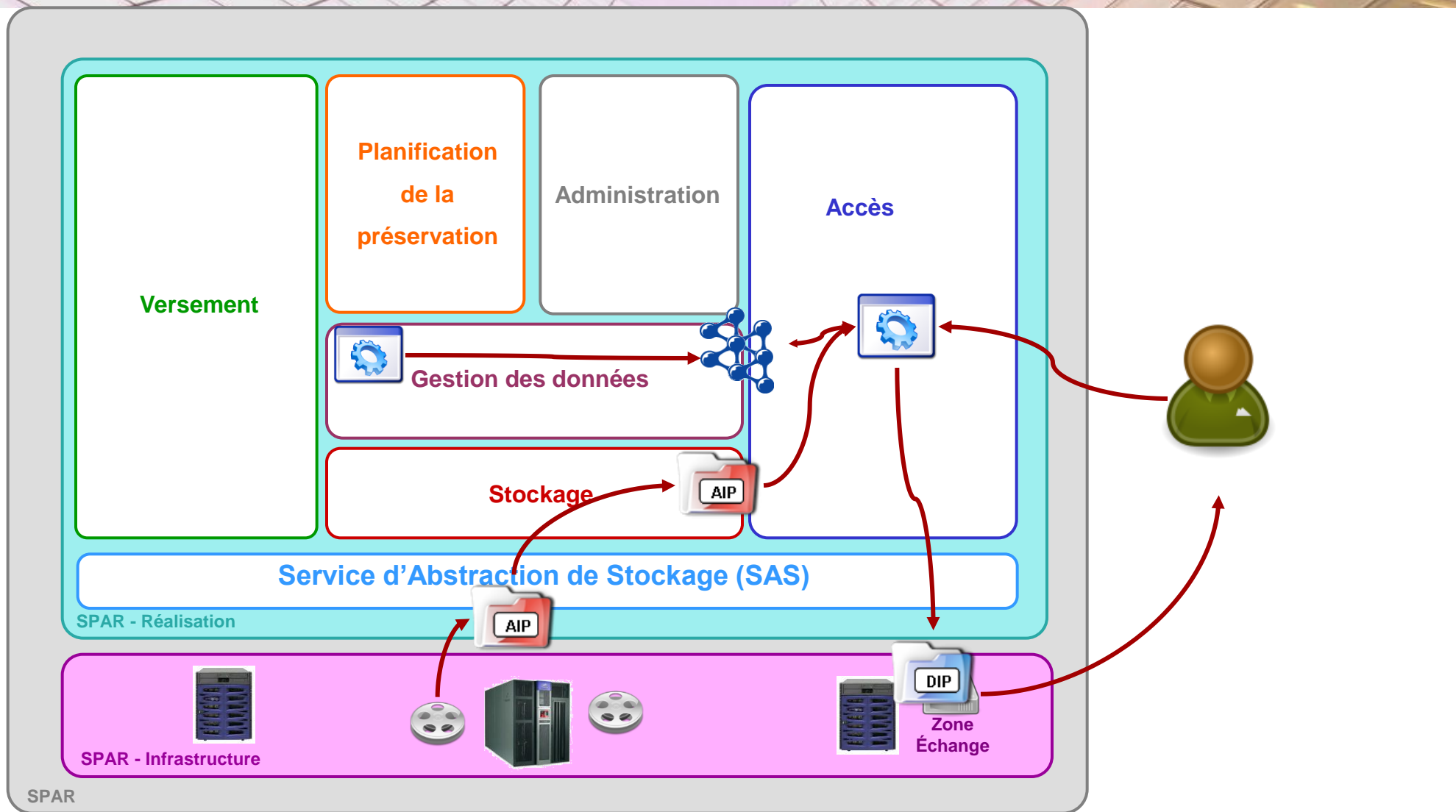
- Une copie bande sur site principal et une copie bande sur le site de secours
- Un audit tous les trois mois

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

La diffusion d'un paquet



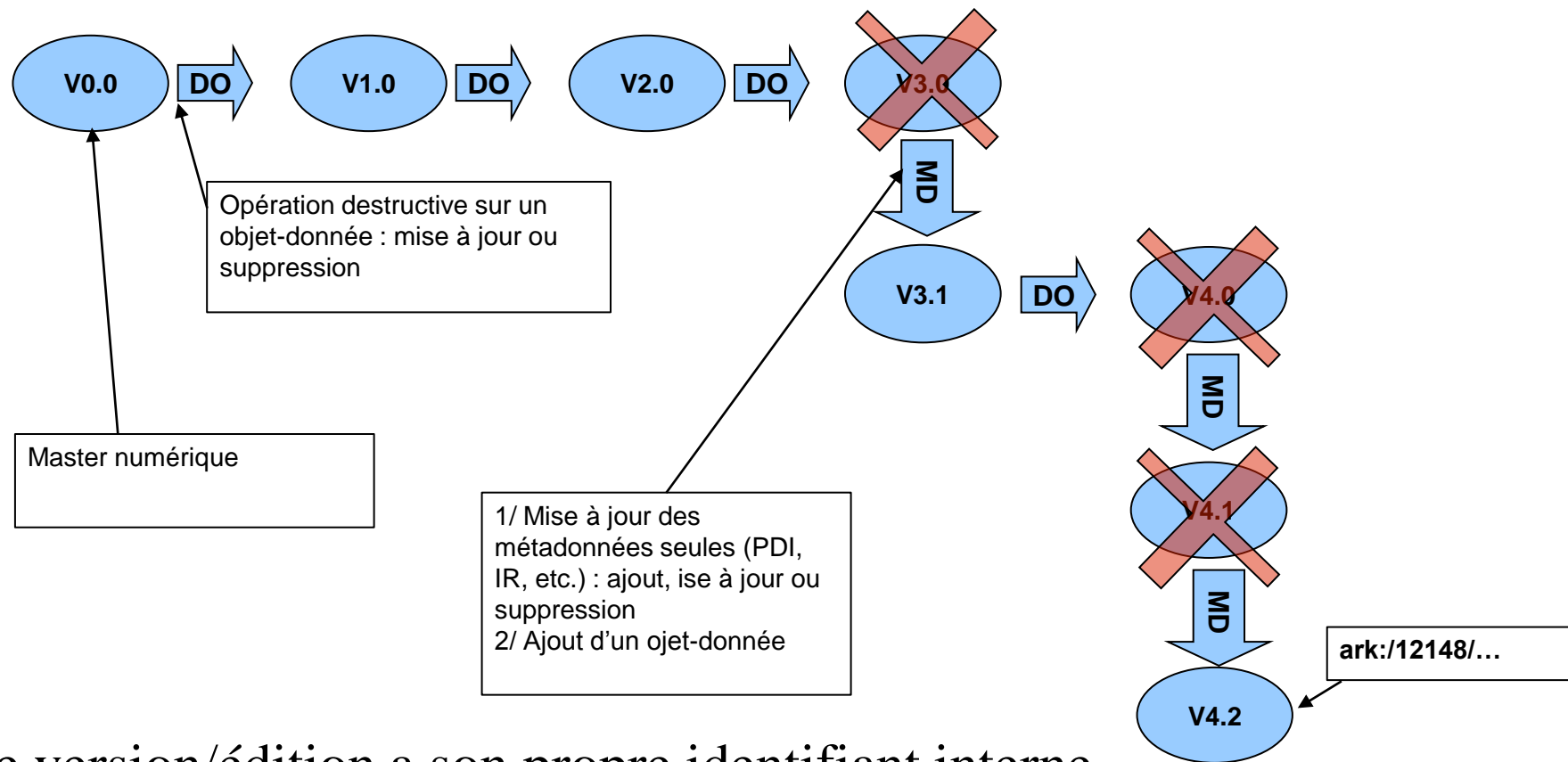
SPAR

Méthodologie

- Les “**filières**” définissent un cadre de contenu pour élaborer la politique de préservation
- Les “**chaînes**” définissent un cadre technique pour appliquer cette politique
- Les “**Accords de Qualité de Service**” fournissent au système une lecture automatisable de la politique
- Les “**paquets de référence**” auto-documentent le système, décrivent les processus, les acteurs...
- Le “**modèle d'information**” se base sur des standards (METS, PREMIS, DC, textMD, MIX, MPEG7)
- Le “**module Gestion de Données**” et les “**bases d'audit**” permettent de savoir précisément ce qu'il y a dans le magasin, et quelles sont les opérations effectuées sur les contenus

→ **Un système auditable**

Cycle de vie du paquet archivé (AIP)

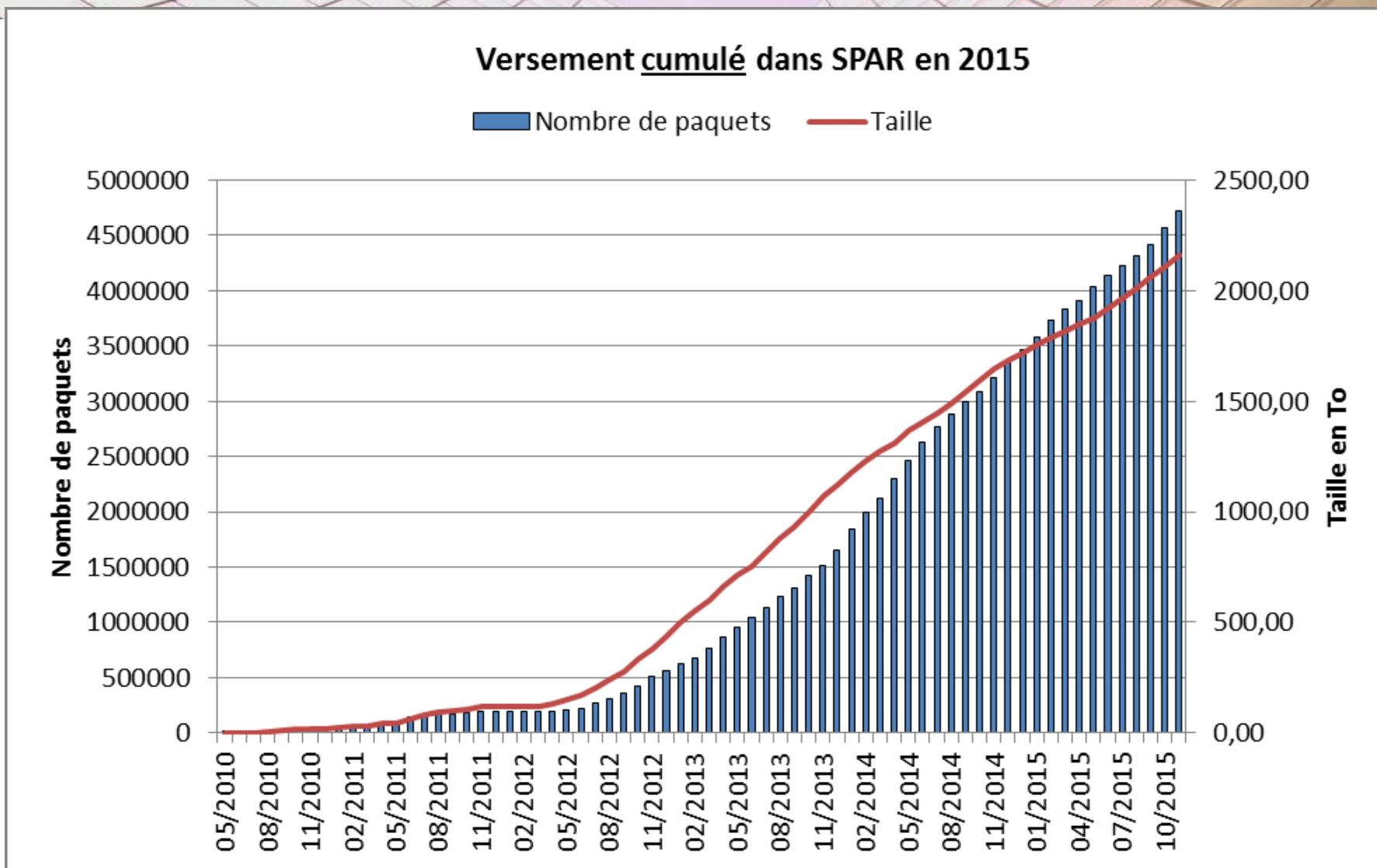


- Chaque version/édition a son propre identifiant interne
- L'identifiant pérenne est unique, au format ark:

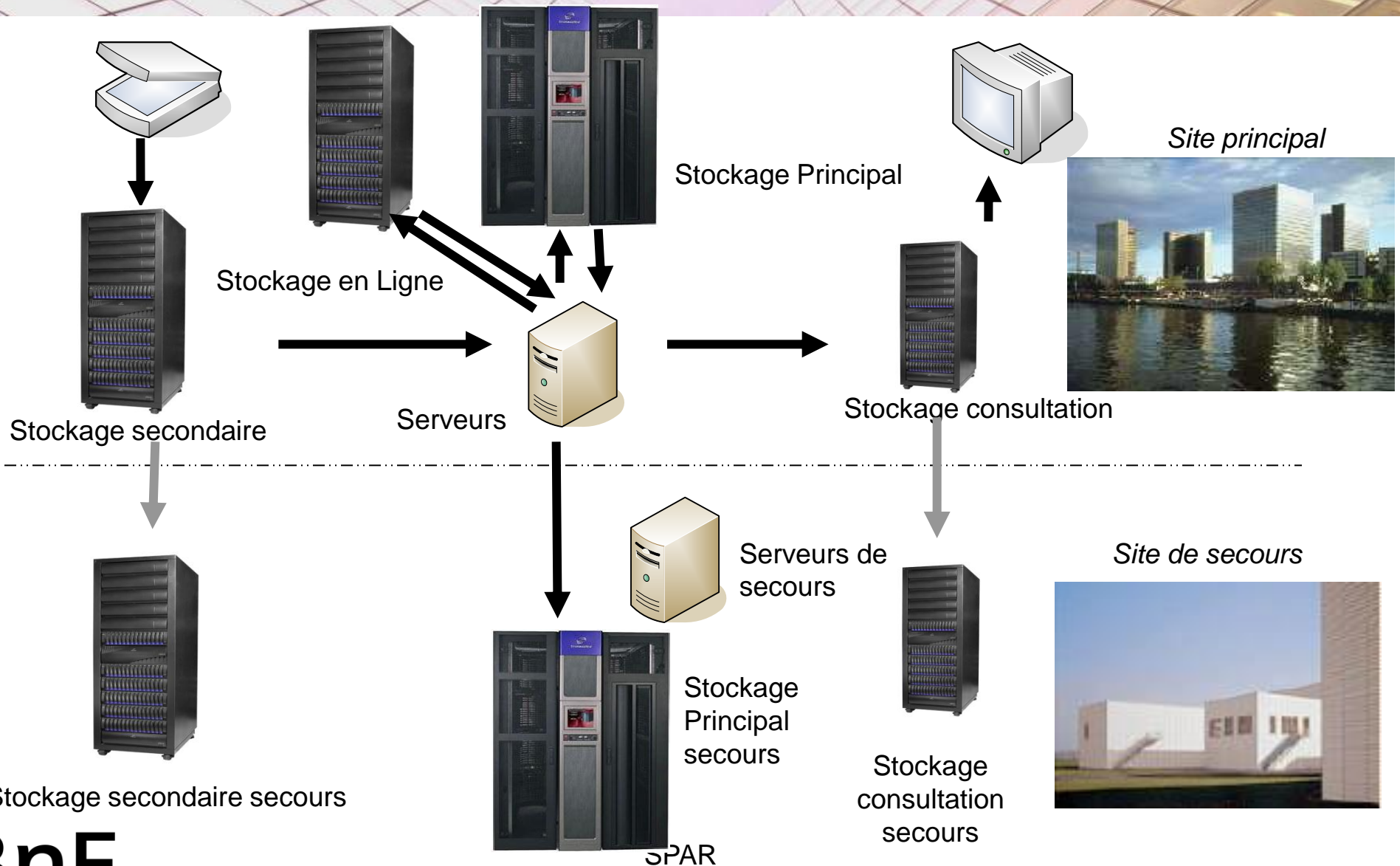
Ordre du jour

- BnF, bref historique et missions
- Préservation numérique à la BnF
- SPAR, le projet
- Modèles de données et flux principaux
- Gérer la masse
- Organisation autour de SPAR

Volumétrie ... toujours plus



Infrastructure matérielle



Infrastructure – Types de matériel

- **Bandes**

- Peu coûteux
- Grande capacité
- Économe en énergie
- Lent (il faut aller chercher la cassette, la charger dans le lecteur...)

⇒ **Pour les documents originaux, et pour ceux dont on ne souhaite pas un accès rapide**

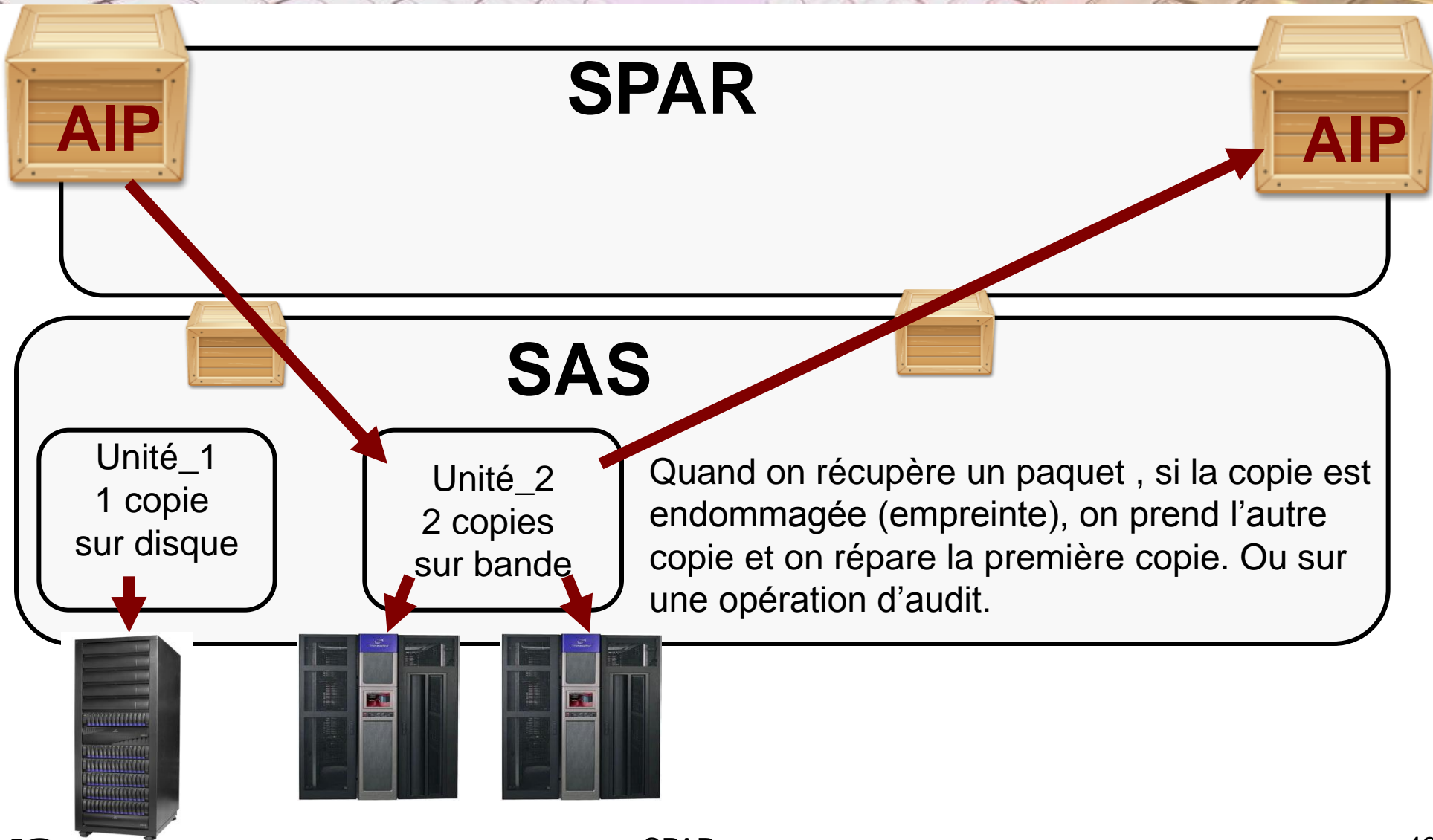
- **Disques**

- Accès rapide
- Plus cher

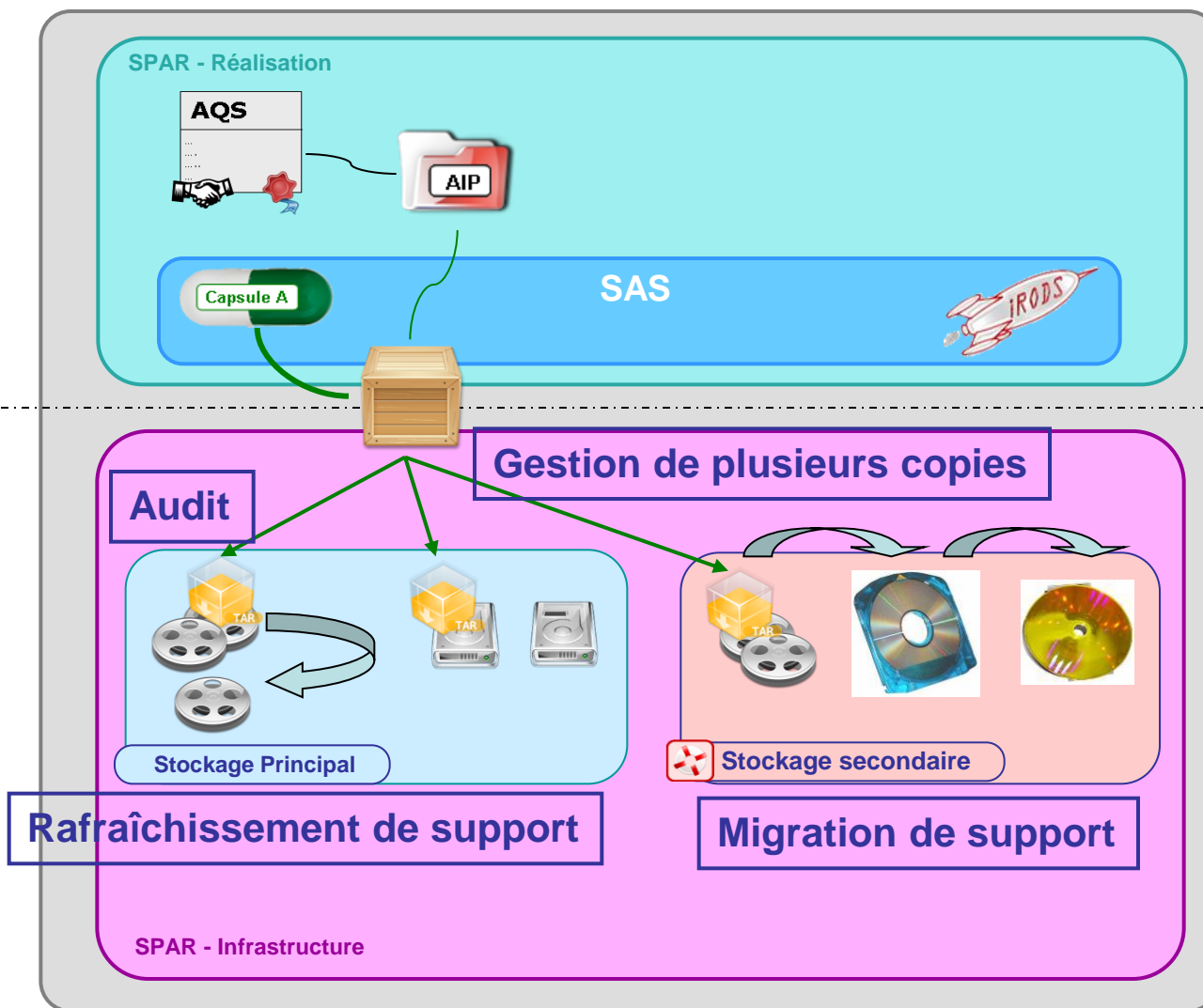
⇒ **Pour les documents en cours de versement,**

⇒ **documents préparés pour la communication rapide**

Stockage



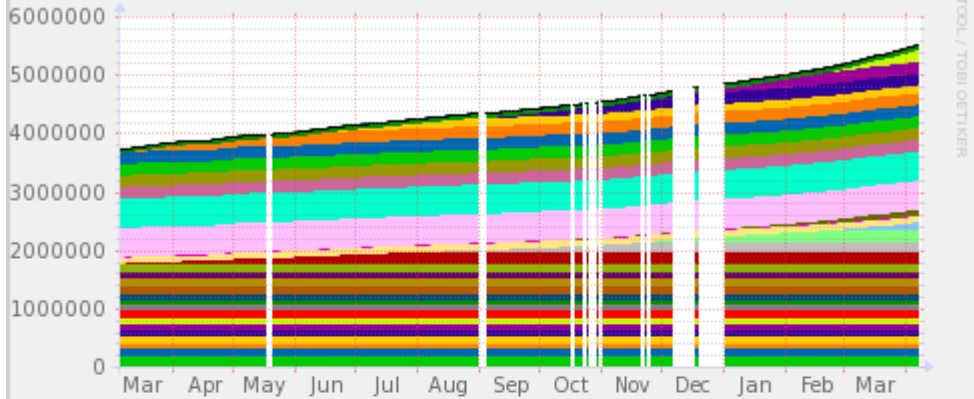
L'abstraction de stockage



Vue Applicative
Vue Infrastructure

Segmentation en Unités de Stockage

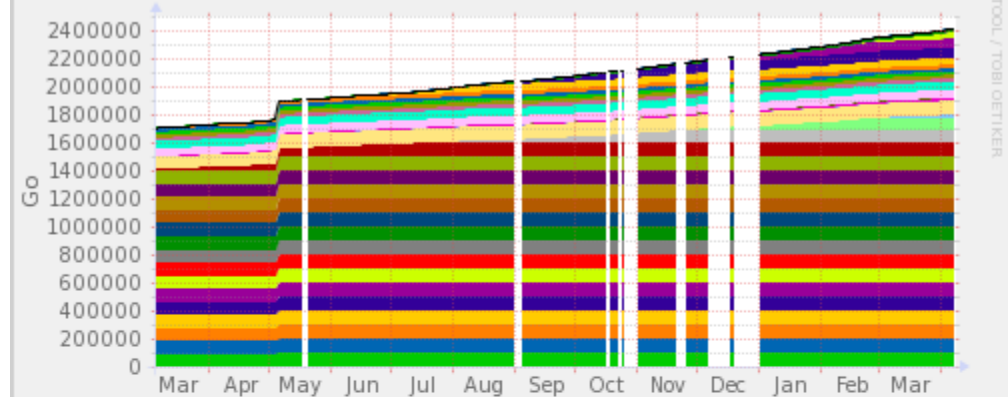
Nombre de AIP par capsule - by year



	Cur:	Min:	Avg:	Max:
capsAIP01	193827.00	193827.00	193827.00	193827.00
capsAIP02	117761.00	117761.00	117761.00	117761.00
capsAIP03	85828.00	85828.00	85828.00	85828.00
capsAIP15	131513.00	131513.00	131513.00	131513.00
capsAIP16	199998.00	23259.00	166254.11	199999.00
capsAIP17	199999.00	1.00	87628.33	199999.00
capsAIP18	199999.00	1.00	98754.49	199999.00
capsAIP19	122237.92	104.00	58948.56	123527.00
capsAIP20	30.00	30.00	30.02	32.00
capsAUD01	110840.00	96834.00	109272.30	110840.00
capsAUD02	784.00	2.00	557.33	784.00
capsCONSB01	10885.00	5970.00	7988.39	10885.00
capsCONSC01	402.00	402.00	402.00	402.00
capsCONSE01	100608.62	158.00	25156.52	101982.00
capsDL01	499999.00	499999.00	499999.00	499999.00
capsDL02	499999.00	499999.00	499999.00	499999.00
capsDL12	12261.36	1908.00	8861.88	15553.00
capsDLAGF01	77.00	77.00	77.00	77.00
capsTA01	88941.05	64800.00	82237.61	89007.00
capsTA02	235.00	5.00	151.58	235.00
Total	5526452.95	3716500.00	4455588.16	5532473.00

Last update: Thu Apr 7 06:11:38 2016
Munin 2.0.17

Occupation des capsules - by year



	Cur:	Min:	Avg:	Max:
capsAIP01	99999.99	93132.25	98839.00	99999.99
capsAIP02	100000.00	93132.26	98839.01	100000.00
capsAIP03	99999.97	93132.23	98838.98	99999.97
capsAIP15	99999.99	93132.25	98839.00	99999.99
capsAIP16	98748.10	17011.01	83572.97	98748.52
capsAIP17	91234.47	0.00	40402.40	91234.47
capsAIP18	82507.61	0.00	39517.82	82507.61
capsAIP19	27402.43	15.03	15619.21	27438.36
capsAIP20	4.13	4.13	4.13	4.40
capsAUD01	100003.22	83130.97	97694.77	100003.22
capsAUD02	1071.77	2.96	781.28	1071.77
capsCONSB01	14979.43	6659.04	10627.61	14979.43
capsCONSC01	26.34	26.34	26.34	26.34
capsCONSE01	4165.23	46.26	2088.87	4184.19
capsDL01	57333.23	53395.73	56667.59	57333.23
capsDL02	57785.96	53817.37	57115.07	57785.96
capsDL12	3392.35	513.14	2451.71	4296.89
capsDLAGF01	341.29	317.85	337.33	341.29
capsTA01	19931.19	7687.06	13204.00	19947.24
capsTA02	5566.06	49.25	3366.92	5566.06
Total	2416185.96	1704877.87	2057382.52	2417161.44

Last update: Thu Apr 7 06:11:38 2016
Munin 2.0.17

Fonctionnalités ... toujours plus

- D'abord, formats de numérisation : maîtrisés techniquement et fonctionnellement
- Puis, collecte du Web : maîtrise du processus par une équipe dédiée
- Et encore : production administrative, les acquisitions numériques

Formats : classification en catégories

	Catégorie	Description
00	Stocké	Pas d'information technique Préservation du train de bits
01	Identifié	Format identifié grâce à un répertoire de formats Pas de trajectoire de préservation planifiée au niveau de l'établissement
10	Connu	Format identifié, documenté, avec des outils de référence Trajectoire de préservation définie au niveau de l'établissement
11	Maîtrisé	Format identifié Documentation et outils de référence maîtrisés par l'établissement Mise en place d'un suivi du format et définition d'un profil d'utilisation au sein de l'établissement

Intégration continue et tests unitaires

Résultat des tests

0 échecs (±0) , Non passés (±0)

792 tests (±0)				
Module	Echec	(Différence)	Total	(Différence)
fr.bnf.spar:access	0		18	
fr.bnf.spar:administration	0		110	
fr.bnf.spar:commons	0		143	
fr.bnf.spar:datamanagement	0		53	
fr.bnf.spar:datamanagement-client	0		4	
fr.bnf.spar:e-signature	0		3	
fr.bnf.spar:ingest	0		215	-3
fr.bnf.spar:ingest-client	0		3	
fr.bnf.spar:manipulation-service-extension	0		6	
fr.bnf.spar:preingest	0		77	
fr.bnf.spar:preingest-common	0		30	
fr.bnf.spar:preingest-ta	0		14	
fr.bnf.spar:preingestaudb	0		6	
fr.bnf.spar:preingestdlauto	0		13	
fr.bnf.spar:preingestprodadm	0		18	
fr.bnf.spar:storage	0		48	
fr.bnf.spar:storage-client	0		28	
fr.bnf.spar:tool-epubcheck	0		3	+3

au 15/12/2014

Tests fonctionnels 0,1%

Génération custom d'un rapport Junit

```
<testsuites>
  <testsuite>
    <testcase classname="packageTest.dlauto" name="762-metadata-1" time="91.11"/>
    <testcase classname="packageTest.dlauto" name="861-metadata-1" time="62.03"/>
  </testsuite>
  <testsuite>
    <testcase classname="packageTest.ta" name="MS1495" time="5.4"/>
    <testcase classname="packageTest.ta" name="MS1496" time="10.2"/>
  </testsuite>
</testsuites>
```

Tous les tests

Class	Durée	Échec	(diff)	Sauté	(diff)	Total	(diff)
/AUDB/	2 h 6 mn	0		0		13	
/CONSA/	1 h 15 mn	0		0		22	
/DLAUTO/	2 j 15 h	0		0		31	
/PRODADM/	1 h 25 mn	11	+7	0		38	
/TA/	5 h 31 mn	4		0		27	

Régressions

packageTest./PRODADM/.BnF-ADM-2008-039468-01 (from (report01pourcent.xml))

En échec depuis 1 build (Depuis [#256](#))

[A duré 59 s.](#)

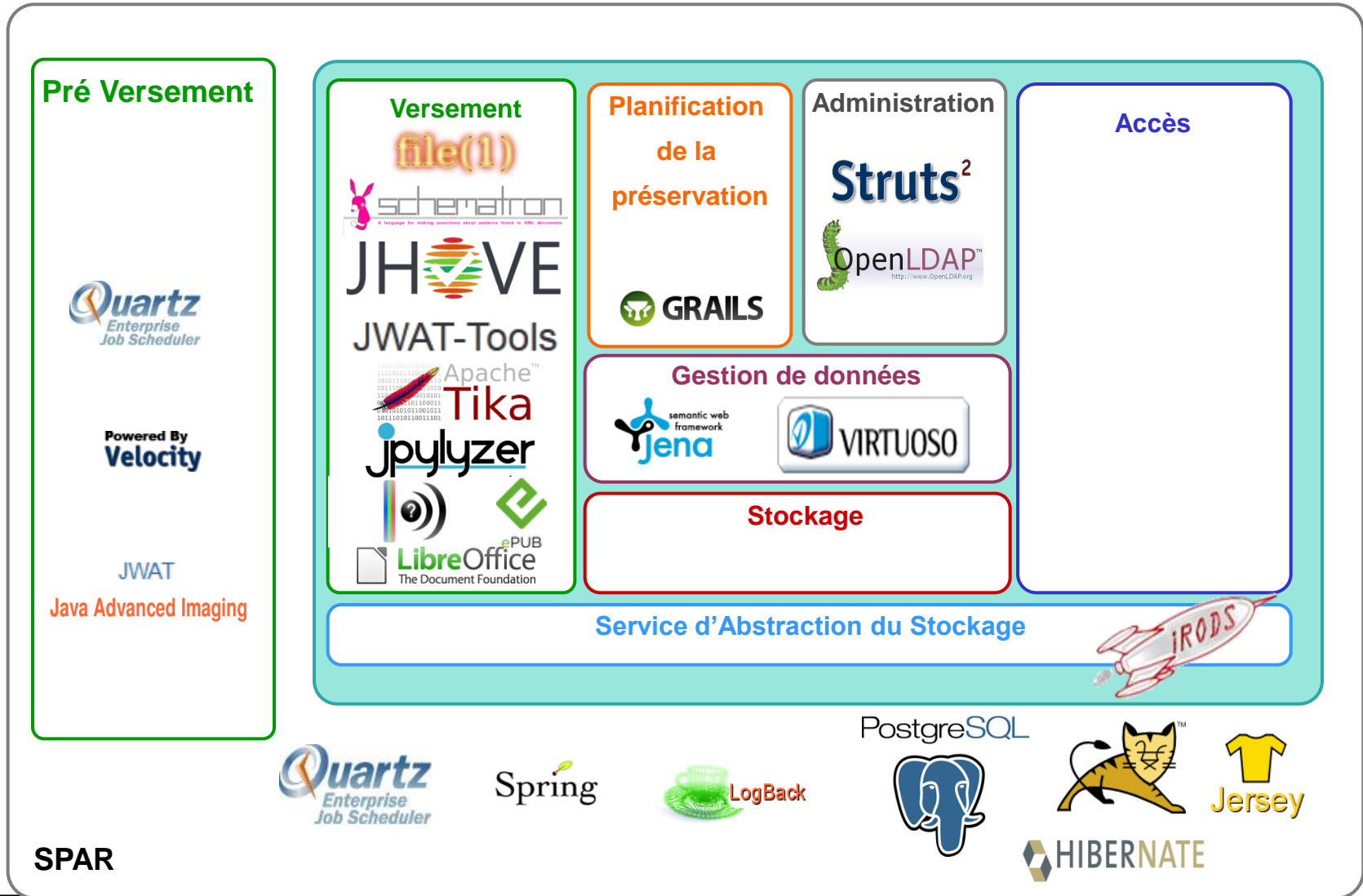
Pile d'exécution

```
Status differ : generated is nok but reference is ok (generated : /partage/spar-
infra/scripts/01pour100/infos_all_package//PRODADM/export_jeu04_majfolder1/Doc_BnF-
ADM-2008-039468-01.info, reference : /partage/spar-infra/scripts/01pour100//infos_references
//PRODADM/export_jeu04_majfolder1/Doc_BnF-ADM-2008-039468-01.info)
```

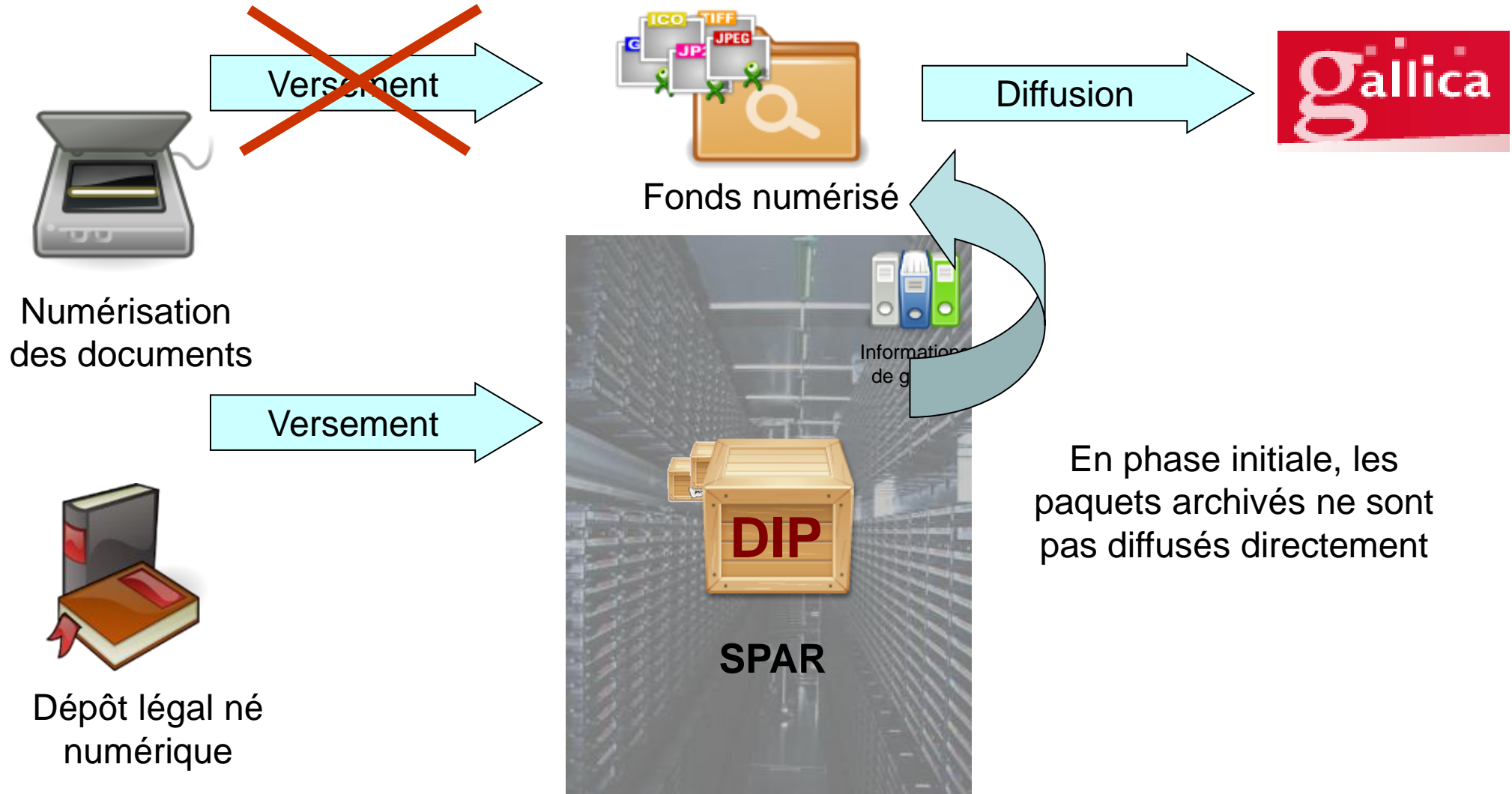
Tests de non-régression 1%

- Réalisé à chaque release
- Page Wiki de suivi
- Jeu de données 1% couvrant toutes les chaînes
 - Analyse initiale à la main
 - Conservation des résultats précédents
 - Passage en automatique
 - « Anonymisation » des résultats: date, uuid, ark
 - Comparaison semi-automatique

Utilisation de briques OpenSource



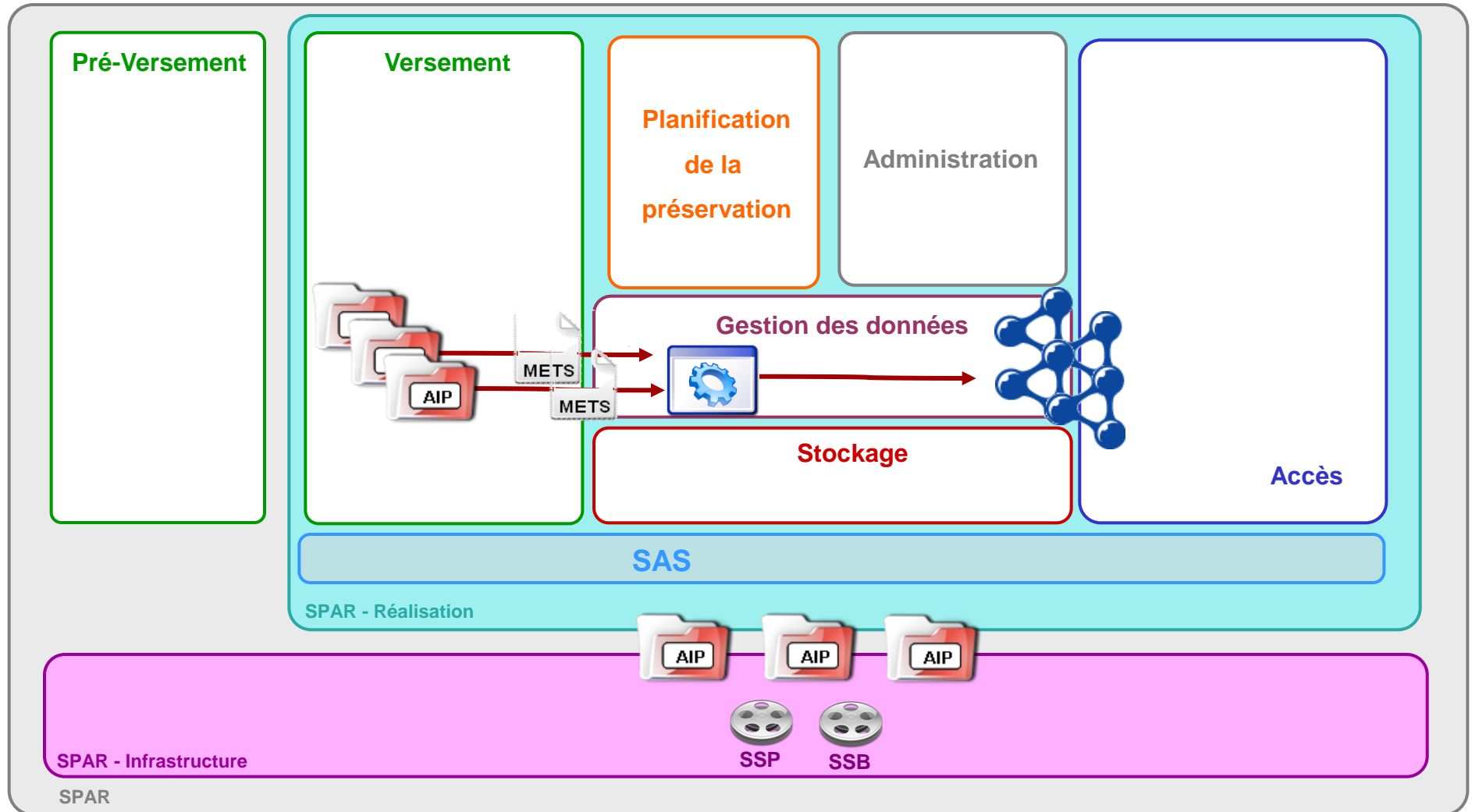
Du schéma en Y vers une archivage initial



Recherche... toujours plus

- Le besoin est une très large gamme de requêtes possibles –non connues à l'avance
- Ces requêtes peuvent porter sur potentiellement toutes les métadonnées
 - Avec des critères multiples
 - Et des attendus variables
- Solution retenue : Web sémantique

Gestion de données : le catalogue des paquets dans SPAR



Pourquoi passer en RDF ?

- METS :

- l'information est structurée
- mais enfermée dans chaque paquet



- RDF :

- l'information libérée de la notion de paquet
- tout en gardant sa forme et sa structure



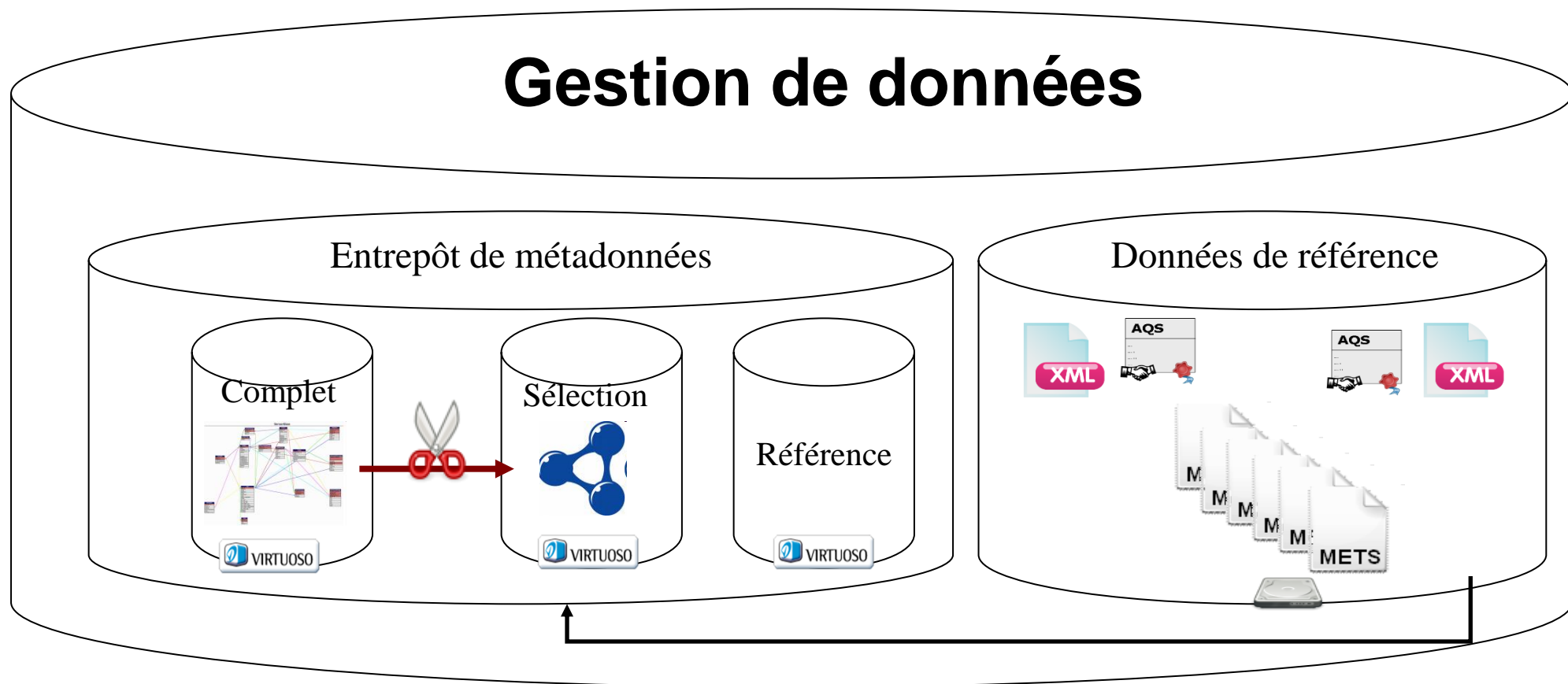
- ... elle est donc plus librement accessible

SPAR

Interroger ces données

- Un modèle pour les décrire : RDF
- Un protocole pour les interroger : SPARQL
 - Toute information présente en RDF peut être
 - Interrogée
 - Comptée
 - Croisée avec d'autres informations
- Permet de récupérer des listes de documents en fonction de caractéristiques

Infrastructure du module Gestion de données



Procédure de reconstruction
SPAR

Entrepôts de métadonnées

- Pour décrire les ~5,5M de paquets
 - Entrepôt complet : 3 162 860 946 triples
 - Entrepôt sélection : 828 701 658 triples
 - Entrepôt référence : 117 070 triples
- ↪ / 4
- Mise en place d'un entrepôt de consultation à destination des experts de préservation

Exemples d'interrogations (1)

- Chargé de collections numériques /
Producteur :
 - Je veux les paquets
 - Où une ou plusieurs pages sont décrites
comme table des matières
 - Mais sans fichier de table des matières en
XML
- Permet de planifier un chantier rétrospectif de
réalisation de tables des matières en XML

Information descriptive

Information de structure

Information de représentation

Exemples d'interrogations (2)

- Contexte de gestion de collection :
 - Pour chaque titre de périodique
 - Je veux dans l'ordre chronologique les dates de publication des fascicules présents dans SPAR
 - Pour chaque monographie en un ou plusieurs volumes
 - Je veux le titre, l'auteur et la date de publication des paquets
 - Entrés dans SPAR le dernier mois
 - Avec le nombre de pages par document

Information de provenance

Information descriptive

Information de structure

Information descriptive

Information de provenance

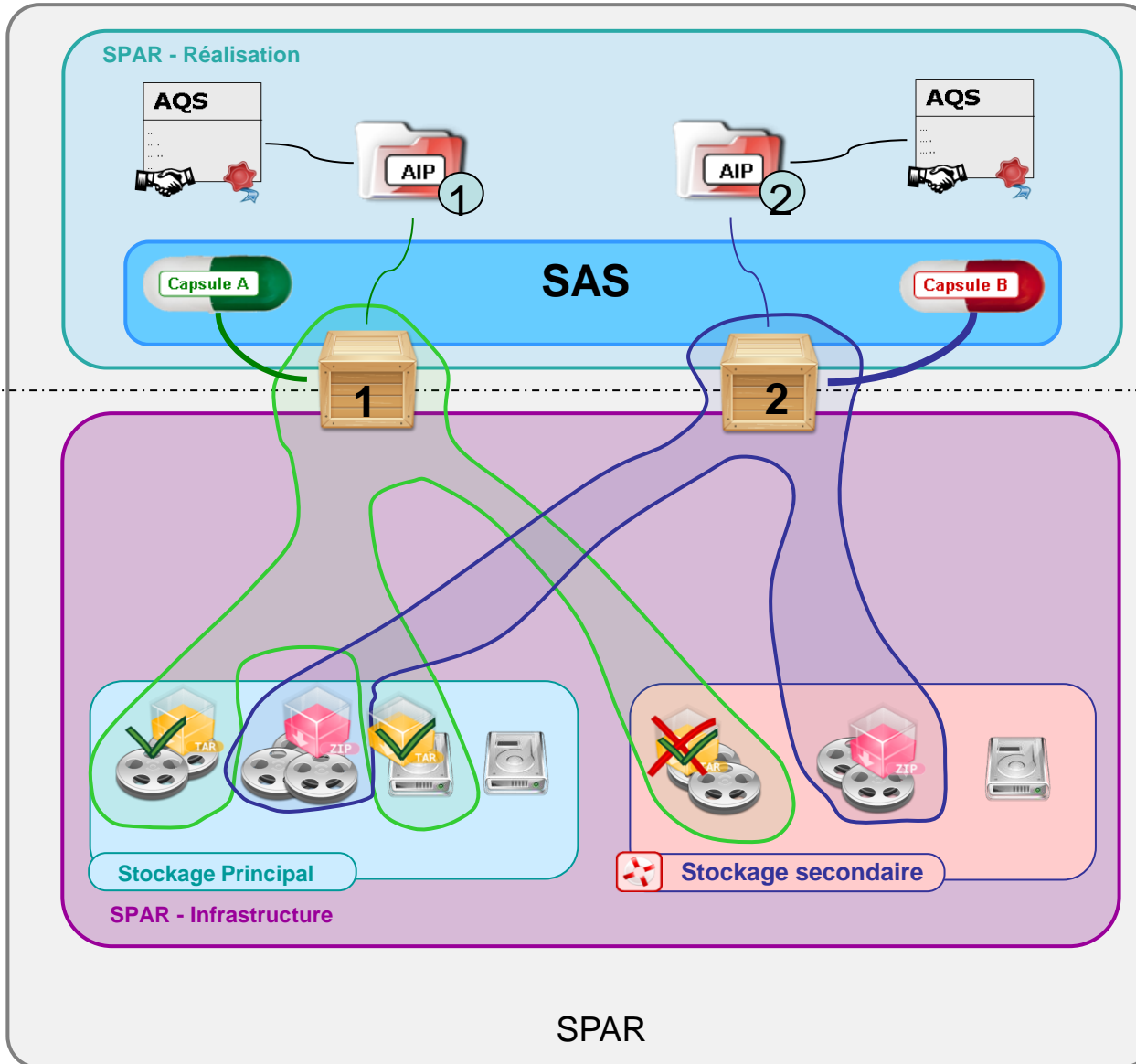
Information de structure

Risques... toujours plus

- Identification des risques
 - Risques projets
 - Risques de sécurité
 - Risques de sûreté
 - Risques technologiques
 - Risques opérationnels
- Elaboration de scénarios au limite

Scénario catastrophe 1

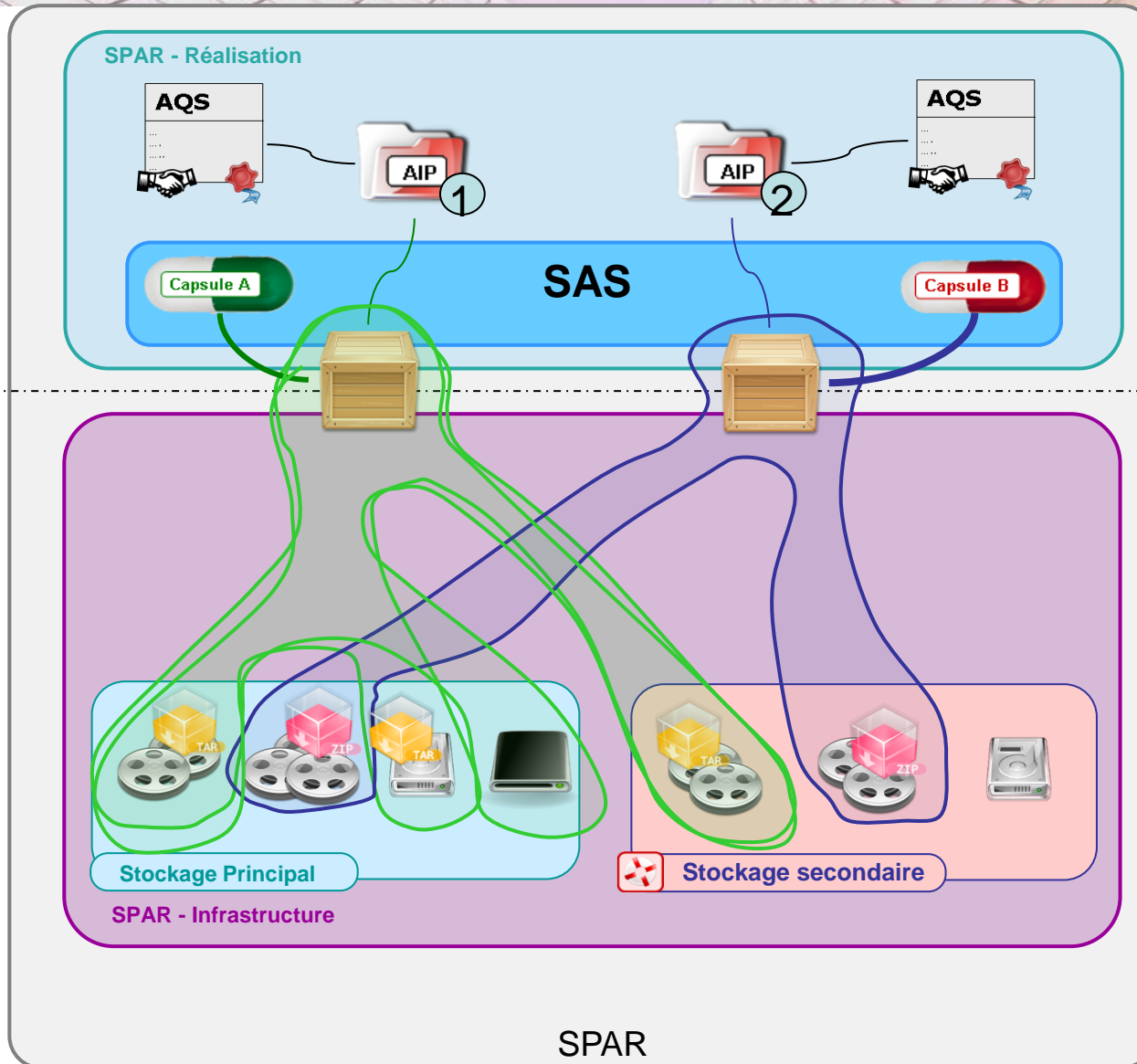
... des copies ne sont plus intègres



Vue Applicative
Vue Infrastructure

Scénario catastrophe 2

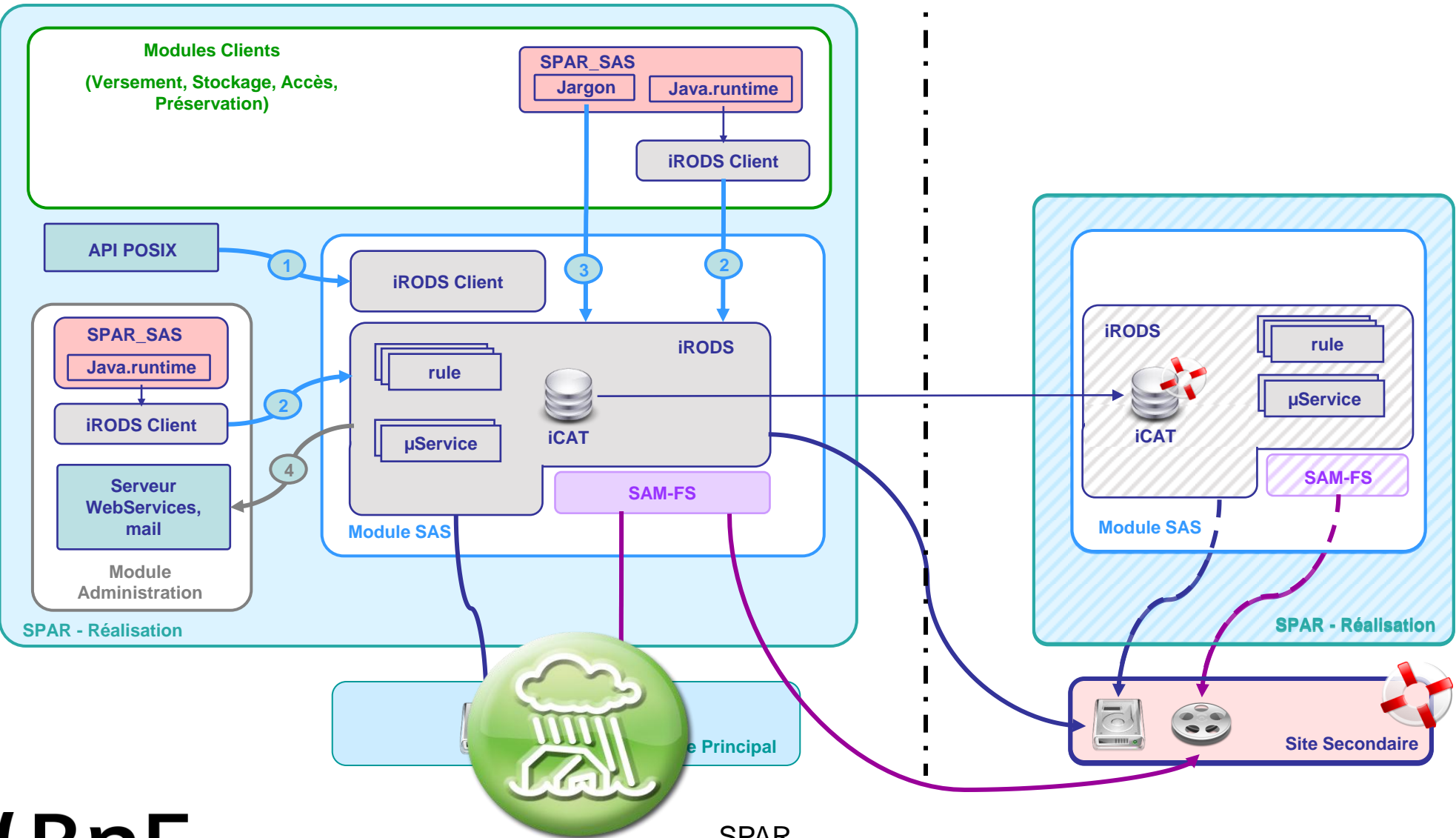
... un élément de stockage devient obsolète



Vue Applicative
Vue Infrastructure

Scénario catastrophe 3

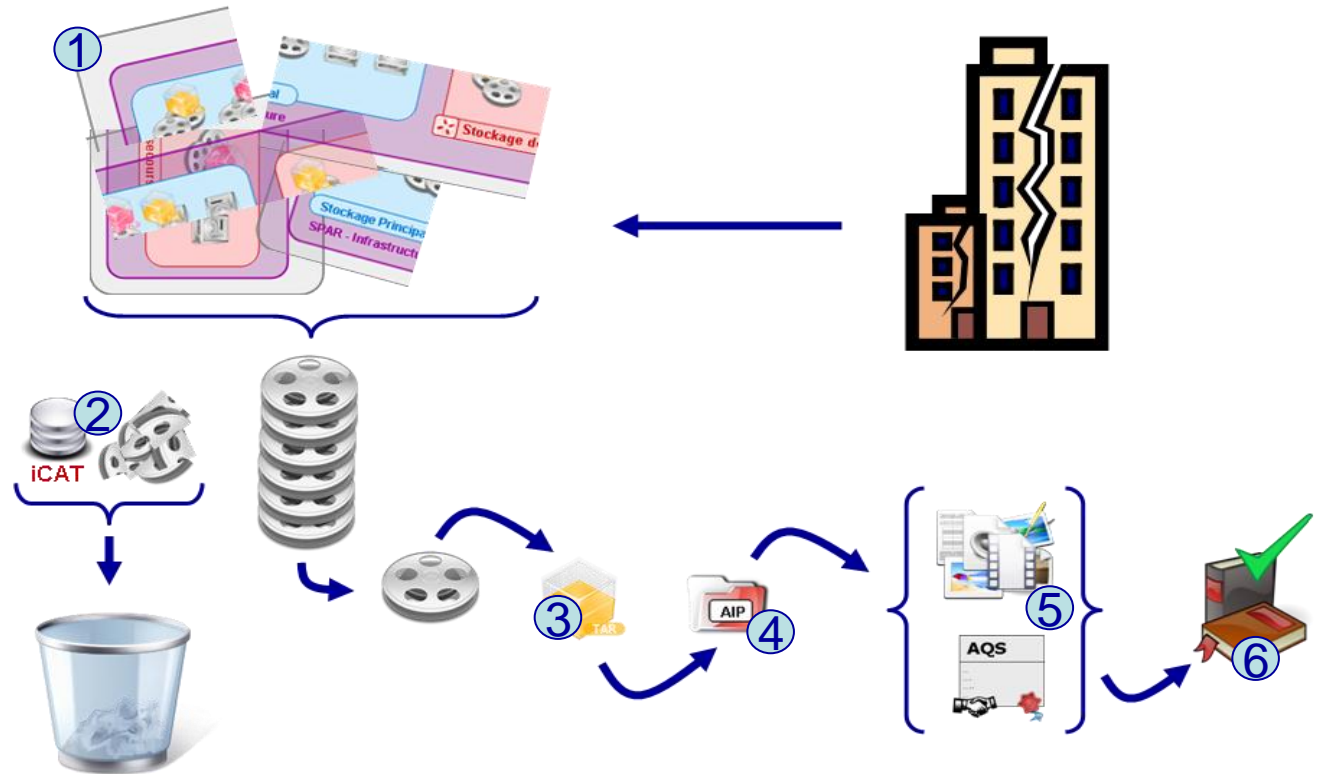
... une inondation dévaste le site principal



Scénario catastrophe 4

... un tremblement de terre détruit les deux sites

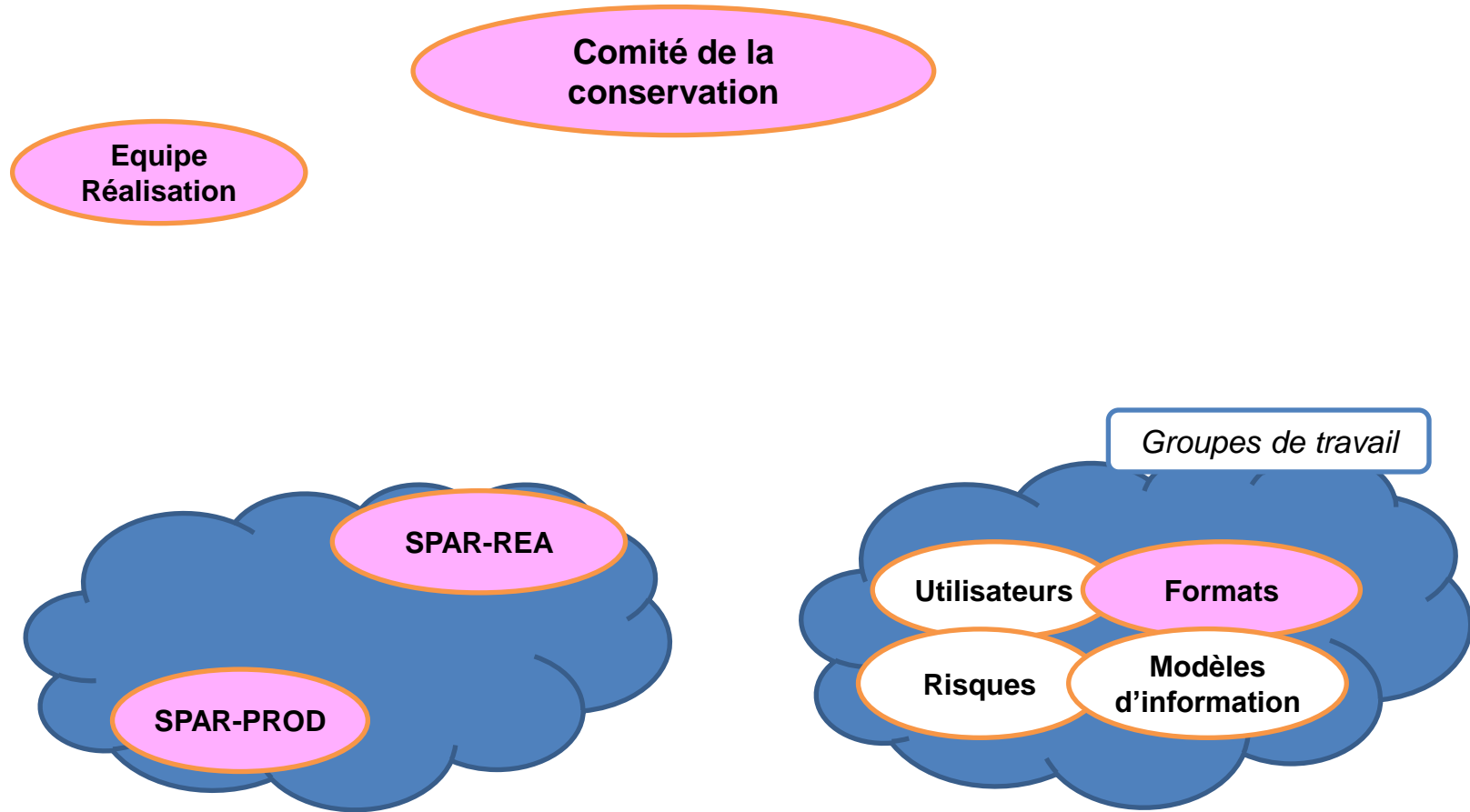
- 1 L'application est inexploitable
- 2 Seules les bandes exploitables sont conservées
- 3 Pour chaque bande, récupération des enregistrements
- 4 Chaque enregistrement matérialise un AIP
- 5 Dans chaque AIP sont présents les objets numériques accompagnés de la caractérisation du paquet.
- 6 Sur la base des objets numériques, des caractérisations et des AQS, il est possible de reconstruire le document numérique.



Ordre du jour

- BnF, bref historique et missions
- Préservation numérique à la BnF
- SPAR, le projet
- Modèles de données et flux principaux
- Gérer la masse
- Organisation autour de SPAR

Organisation début de projet



Évolution des métiers

- Relation renforcée avec la technique
 - données stockées sur l'infrastructure de stockage
 - opérations de migration sous le contrôle d'administrateur système
- Évolution du rôle du DSI : passage d'un fournisseur de services vers un partenaire de la gestion des collections

Interaction avec SPAR : les rôles

Gérer le magasin
numérique



Interaction avec SPAR : les rôles

Apporter l'expertise de préservation : formats, cohérence des données, ingénierie informatique



Interaction avec SPAR : les rôles



Gérer les contenus numériques

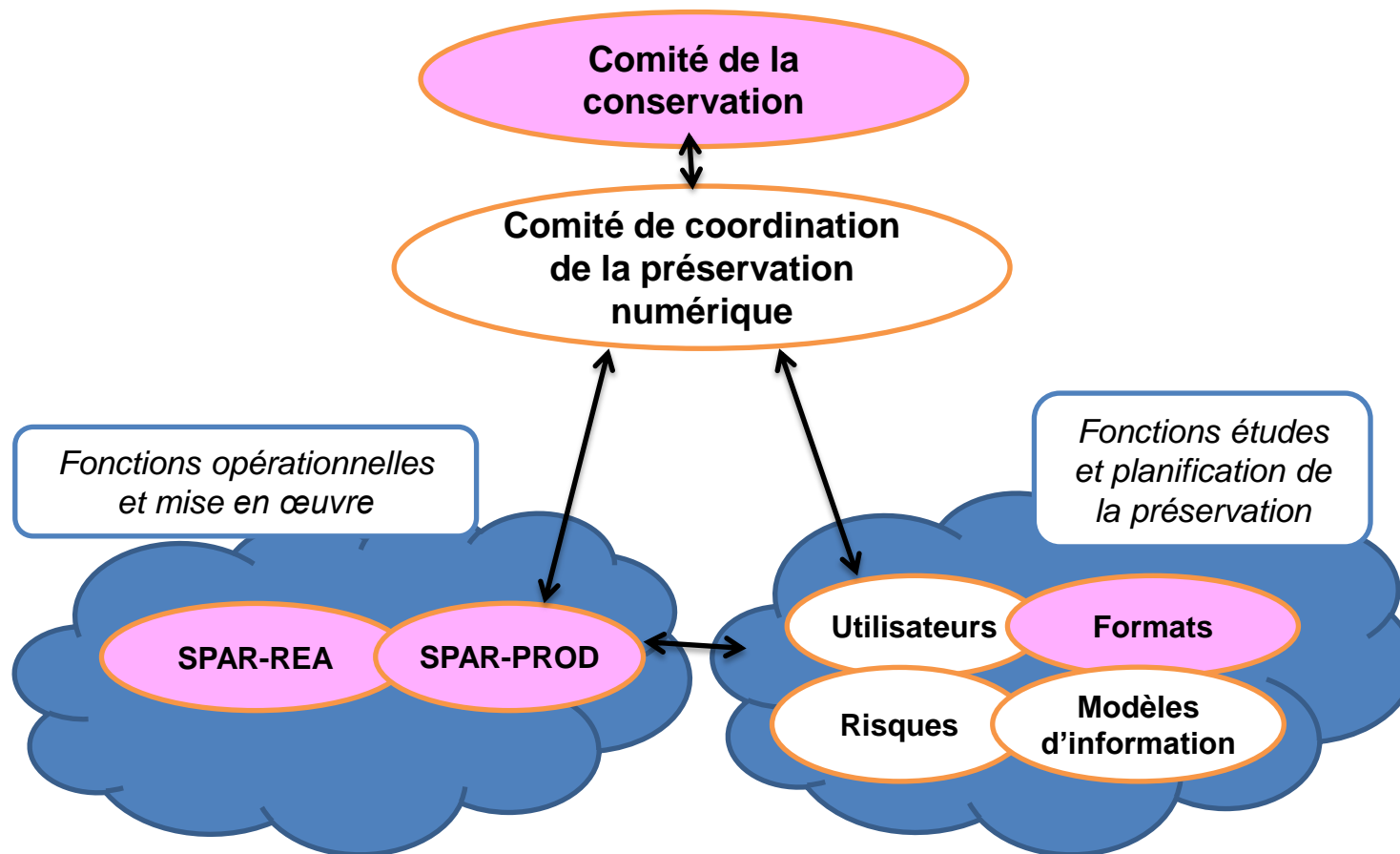


Interaction avec SPAR : les rôles

Mais encore



Organisation de la préservation



Nécessité de veille

- Conventions avec plusieurs pays sur le numérique
- Traductions
 - Norme OAIS
 - METS : vue d'ensemble & guide
 - PREMIS : vue d'ensemble
- Au niveau national, BnF pôle de compétence sur le numérique :
 - **Formations à destination du Réseau des pôles associés et des professionnels de l'information**
 - **Élaboration de guides pratiques et de recommandation**

Collaboration aux groupes de travail sur la conservation du numérique

- Niveau national :
 - Groupe PIN (Pérennité Information Numérique)
 - AFNOR/CN 171
 - SNB (Schéma Numérique des Bibliothèques)
 - BSN (Bibliothèque Scientifique Numérique)
- Niveau international :
 - IFLA
 - Consortium IIPC : collecte et préservation du Web
 - PASIG : Preservation & Archiving Special Interest Group
 - OpenPreservation Foundation (OPF)
 - Keeping Emulation Environments Portable (KEEP)
 - Unified Digital Format Registry (UDFR)
 - METS Editorial Board
 - PREMIS Editorial Committee

Implications dans les outils de base

- BnF est un miroir technique pour les identifiants pérennes ark (Archival Resource Key) depuis 2010
- Participations aux logiciels (contributions, roadmap)
 - **iRods**
 - **Jhove, jpylyzer**
 - **JWAT**
 - **Epubcheck**
- Contrats commerciaux
 - **MedialInfo**
 - **Virtuoso**

SPAR : Un système opérationnel

toc : 1 FILE(S)

master : 511 FILE(S)

ocr : 511 FILE(S)

[Afficher le manifest](#)

Nom : ark-12148-bpt6k102002g.version0.release1.tar

Catégorie : AIP

Statut : OK

MD5 : 51e631d3058e7abad55c2c3206ef12d2

Informations de stockage :

Copie	Élément de stockage	Chemin	volume ID	MD5
copie 1	sparresc	/stockage/irods/sparresc/AIP/9968/6414/152/ark-12148-bpt6k102002g.version0.release1.tar		51e631d3058e7abad55c2c3206ef12d2

Processus associés au paquet : [Afficher toutes les actions](#)

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

Jeton	Titre	Statut	Date de début	Date de fin	
bd7a7870-dc62-11df-ba89-00144f68e1cc	ING_1	completed	2010/10/20 17:57:28	2010/10/20 18:23:19	Afficher les actions
5e19c900-dc65-11df-af65-00144f68e1cc	STO_1	completed	2010/10/20 18:16:17	2010/10/20 18:17:25	Afficher les actions
b0a9f460-dc65-11df-a2b3-00144f68e1cc	DM_2	completed	2010/10/20 18:18:35	2010/10/20 18:23:17	Afficher les actions
a4ce33c0-dc6c-11df-88ee-00144f68e7e0	ACC_1	completed	2010/10/20 19:08:22	2010/10/20 19:09:22	Afficher les actions
a7038a50-dc6c-11df-a73e-00144f68e7e0	STO_2	completed	2010/10/20 19:08:26	2010/10/20 19:09:13	Afficher les actions
2d221ac0-dc6d-11df-ba89-00144f68e1cc	ING_1	completed	2010/10/20 19:12:11	2010/10/20 19:24:15	Afficher les actions
64adc520-dc6d-11df-90f2-00144f68e7e0	STO_2	completed	2010/10/20 19:13:44	2010/10/20 19:14:00	Afficher les actions
b03aefd0-dc6e-11df-90f2-00144f68e7e0	STO_1	completed	2010/10/20 19:23:00	2010/10/20 19:23:15	Afficher les actions
bab98d40-dc6e-11df-85be-00144f68e7e0	DM_2	completed	2010/10/20 19:23:18	2010/10/20 19:24:15	Afficher les actions

Salle informatique principale

