



Mutualisation d'infrastructure Migration des archives sur bandes

Thomas Ledoux

Coordinateur de la Production Numérique, DSI, BnF

thomas.ledoux_AT_bnf.fr

Journée PIN du 17 novembre 2022



Ordre du jour

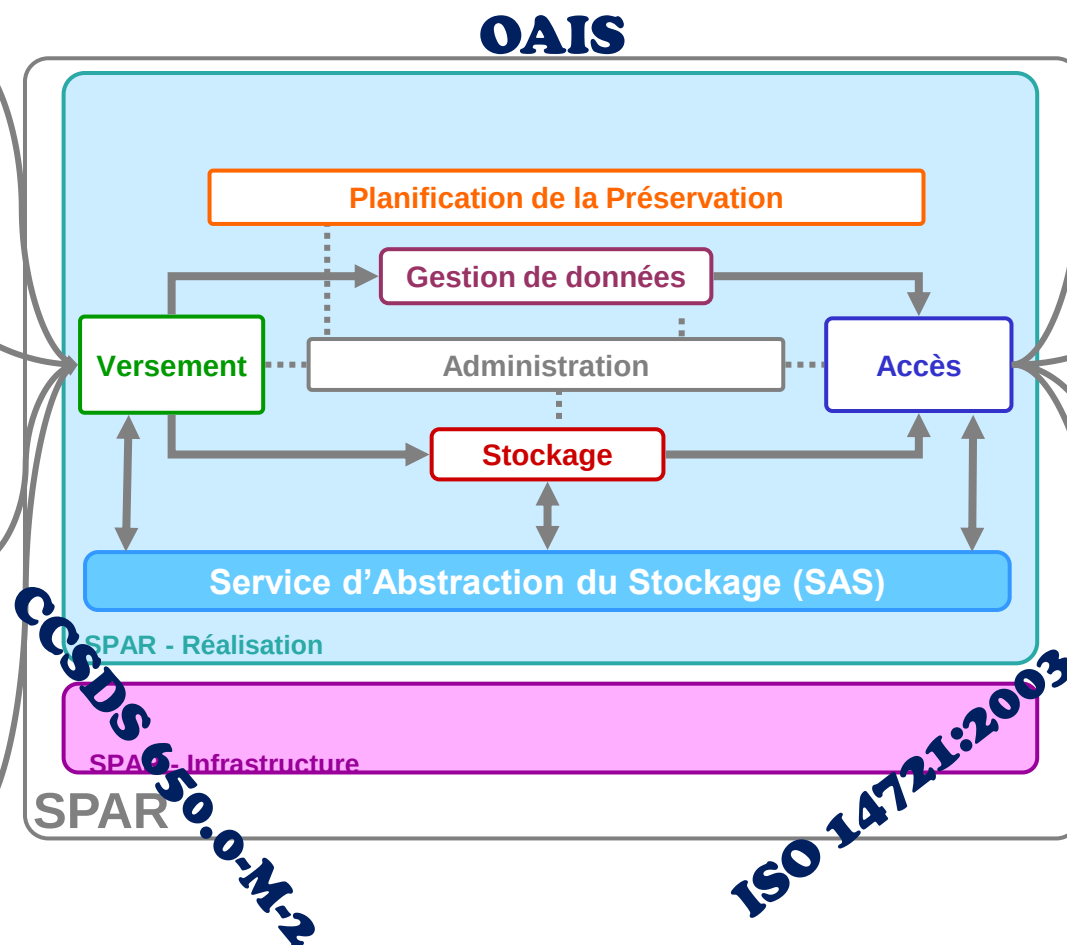
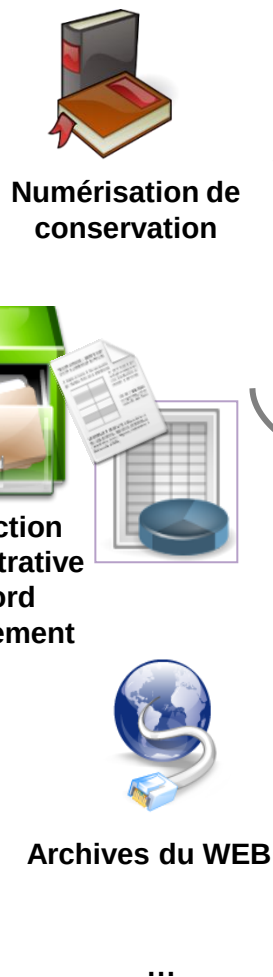
- La fonction stockage dans SPAR
- Migration des bandes au sein de SPAR
- Mutualisation de l'infrastructure



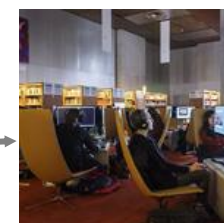
SPAR : stockage

Le Système de Préservation et d'Archivage Réparti (SPAR) de la BnF

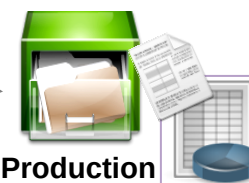
Applications de production de données



Gallica



Postes publics



Production administrative



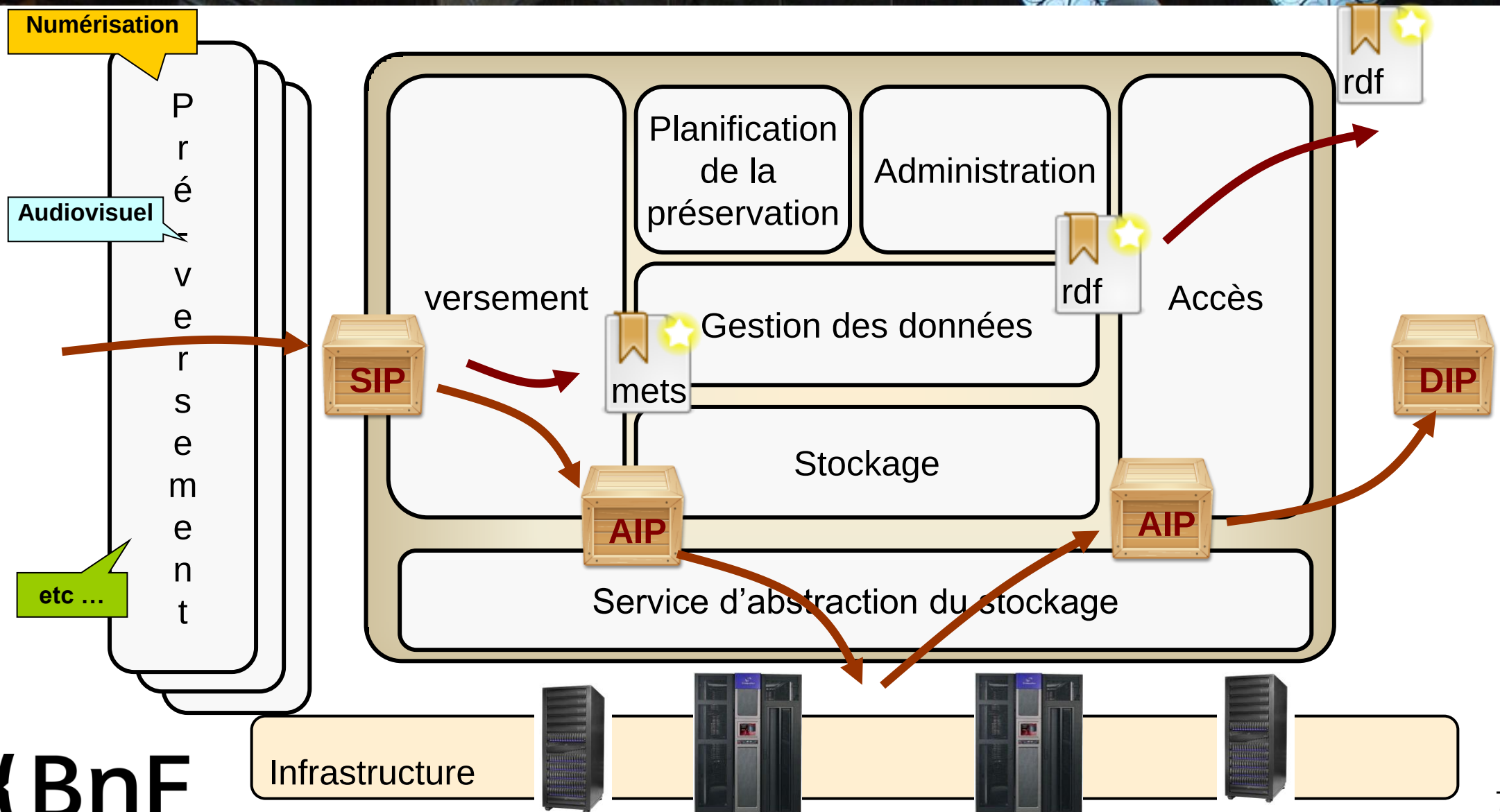
wayback

Applications de diffusion de données

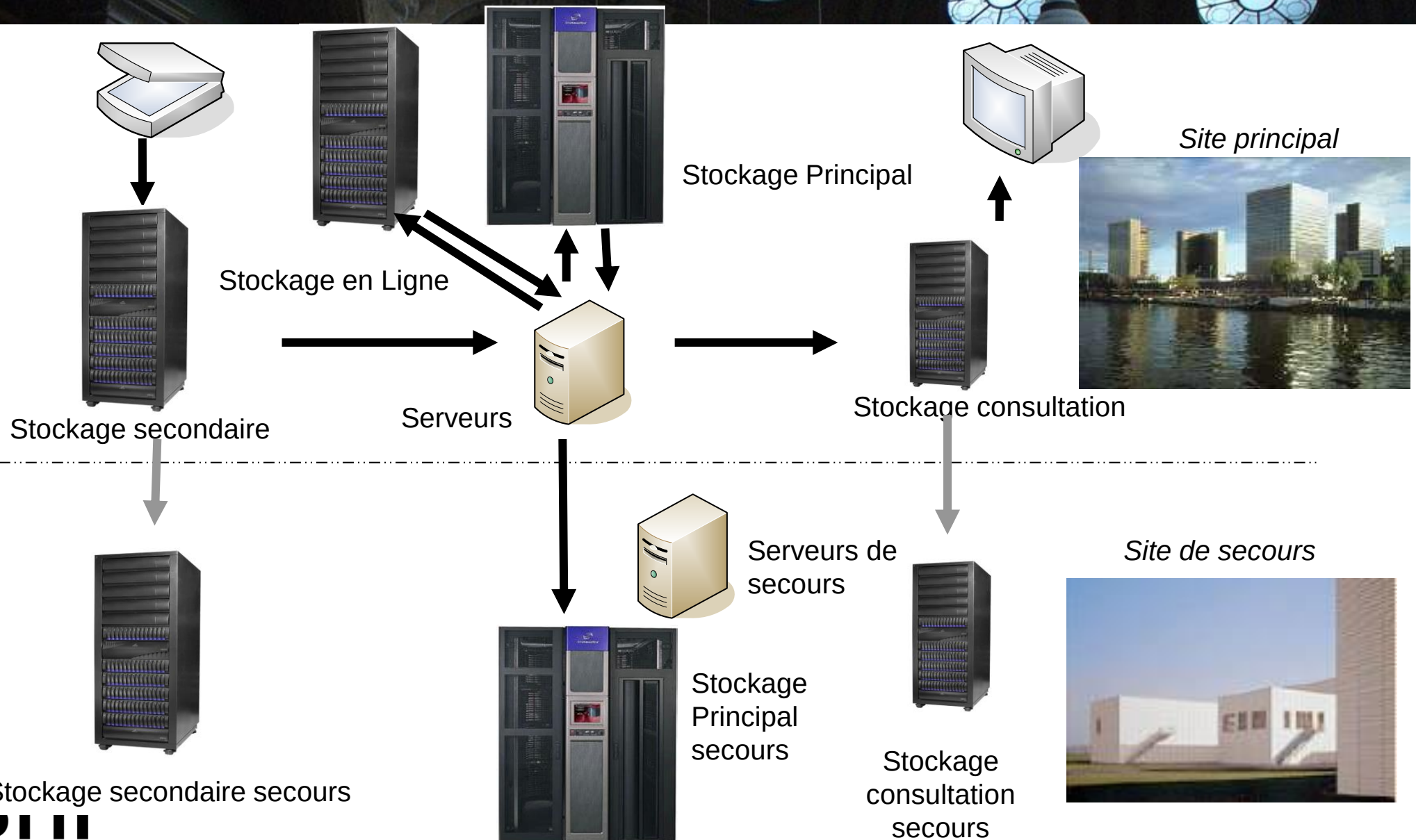
SPAR en chiffres (au 28/10/2022)

- Ouverture de l'exploitation : mai 2010
 - Préservation de la numérisation courante de la BnF
 - Stockage répliqué sur deux sites
- 13 903 488 paquets archivés
- 508 535 064 objet-données (fichiers)
- > 5,2 Po (= 5 207 To) de données brutes
- > 7 milliards de triples (métadonnées élémentaires)
- Des manuscrits très volumineux ont été versés (paquets de taille supérieure à 100 Go)
- Filières disponibles : monographies, périodiques, lots d'image, audio, vidéo, archivage du web, documents bureautiques, objets 3D, tiers archivage

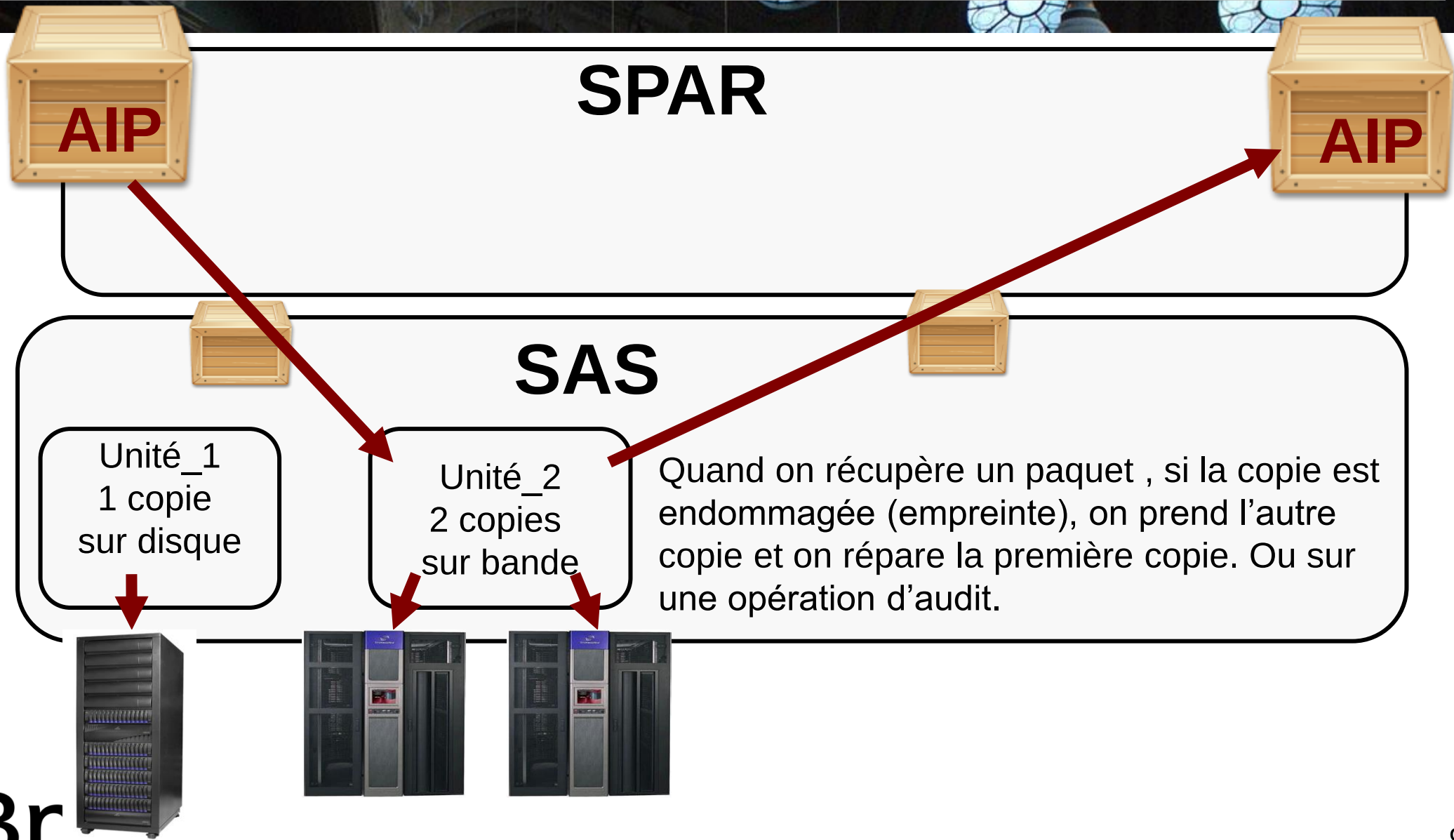
Architecture générale



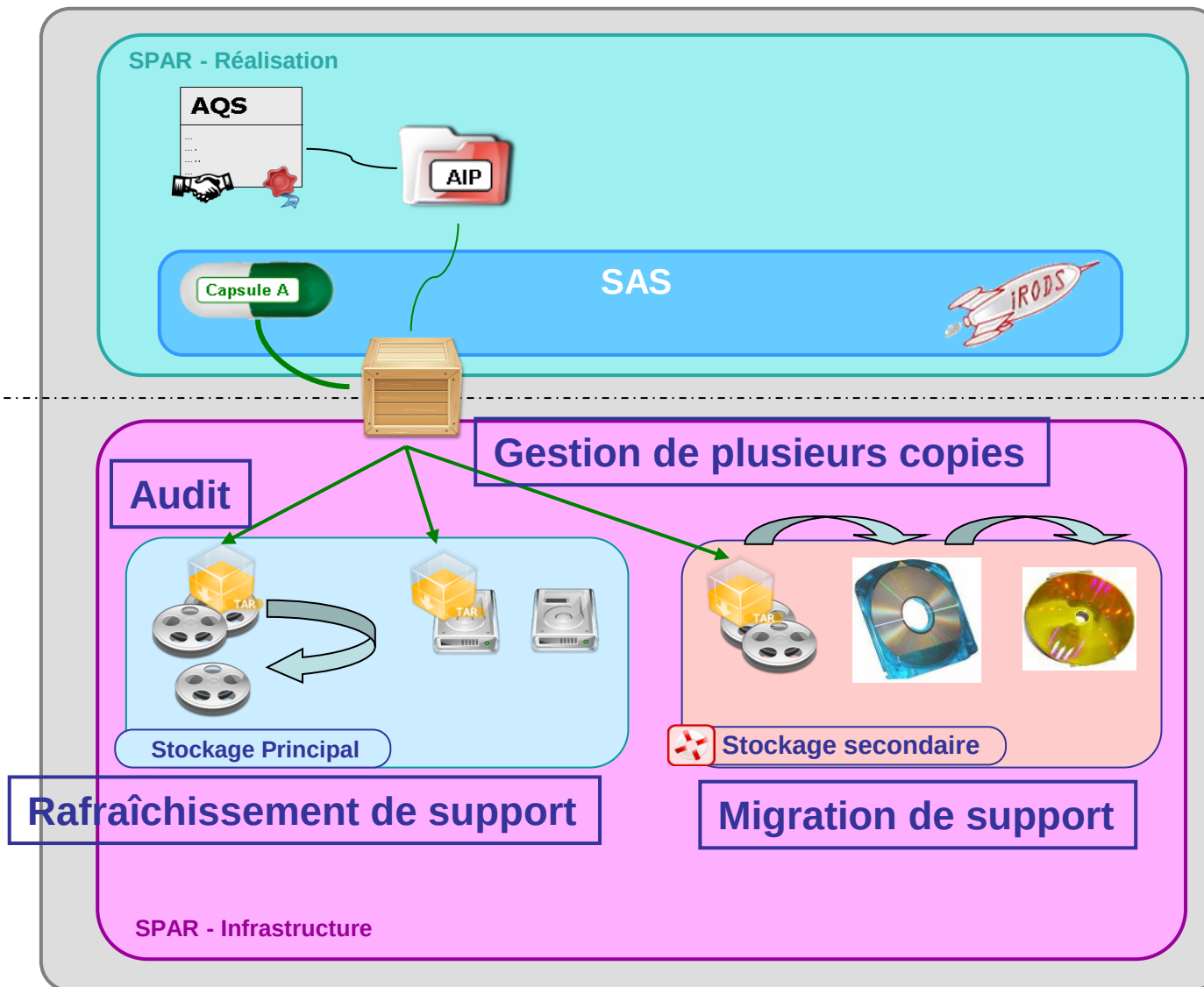
Infrastructure matérielle



Abstraction de Stockage



L'abstraction de stockage



Vue Applicative

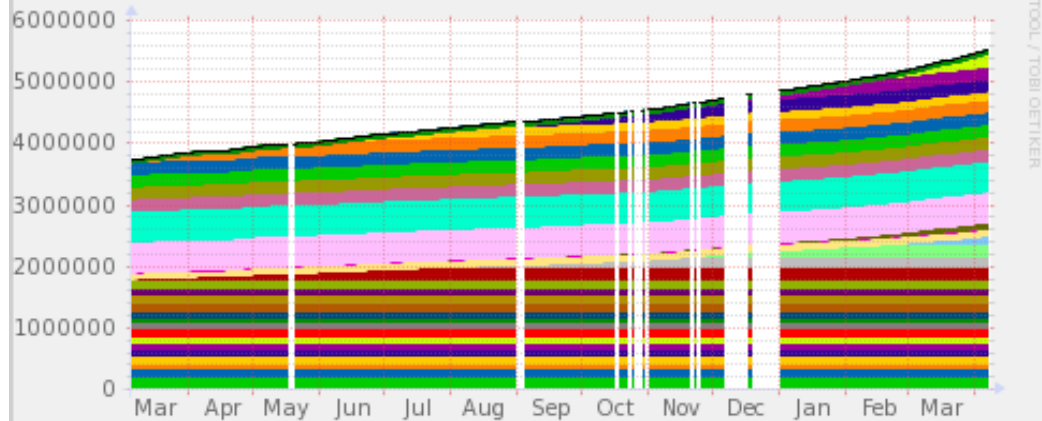
Vue Infrastructure



SPAR : migration

Segmentation en Unités de Stockage

Nombre de AIP par capsule - by year

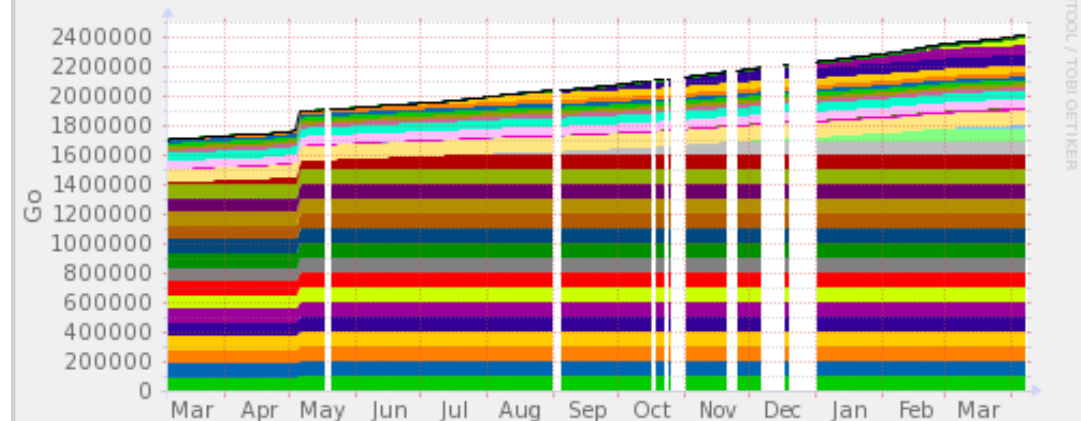


	Cur:	Min:	Avg:	Max:
capsAIP01	193827.00	193827.00	193827.00	193827.00
capsAIP02	117761.00	117761.00	117761.00	117761.00
capsAIP03	85828.00	85828.00	85828.00	85828.00
capsAIP15	131513.00	131513.00	131513.00	131513.00
capsAIP16	199998.00	23259.00	166254.11	199999.00
capsAIP17	199999.00	1.00	87628.33	199999.00
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	445588.16	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

Gestion des capsules

SPAR Réalisation

Connecté(e) en tant que : spar_admin
Déconnexion

Versement

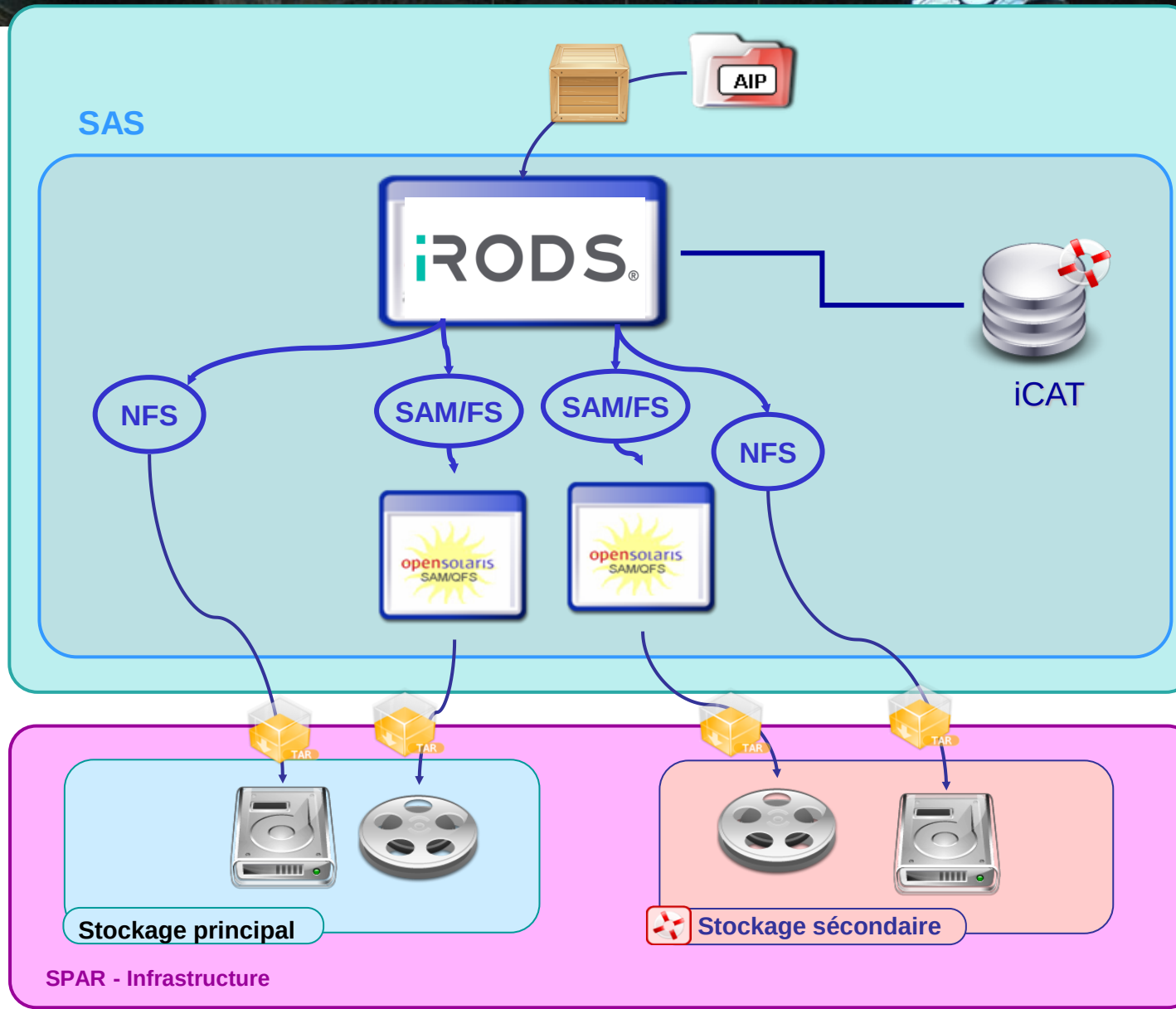
Accès

Administration

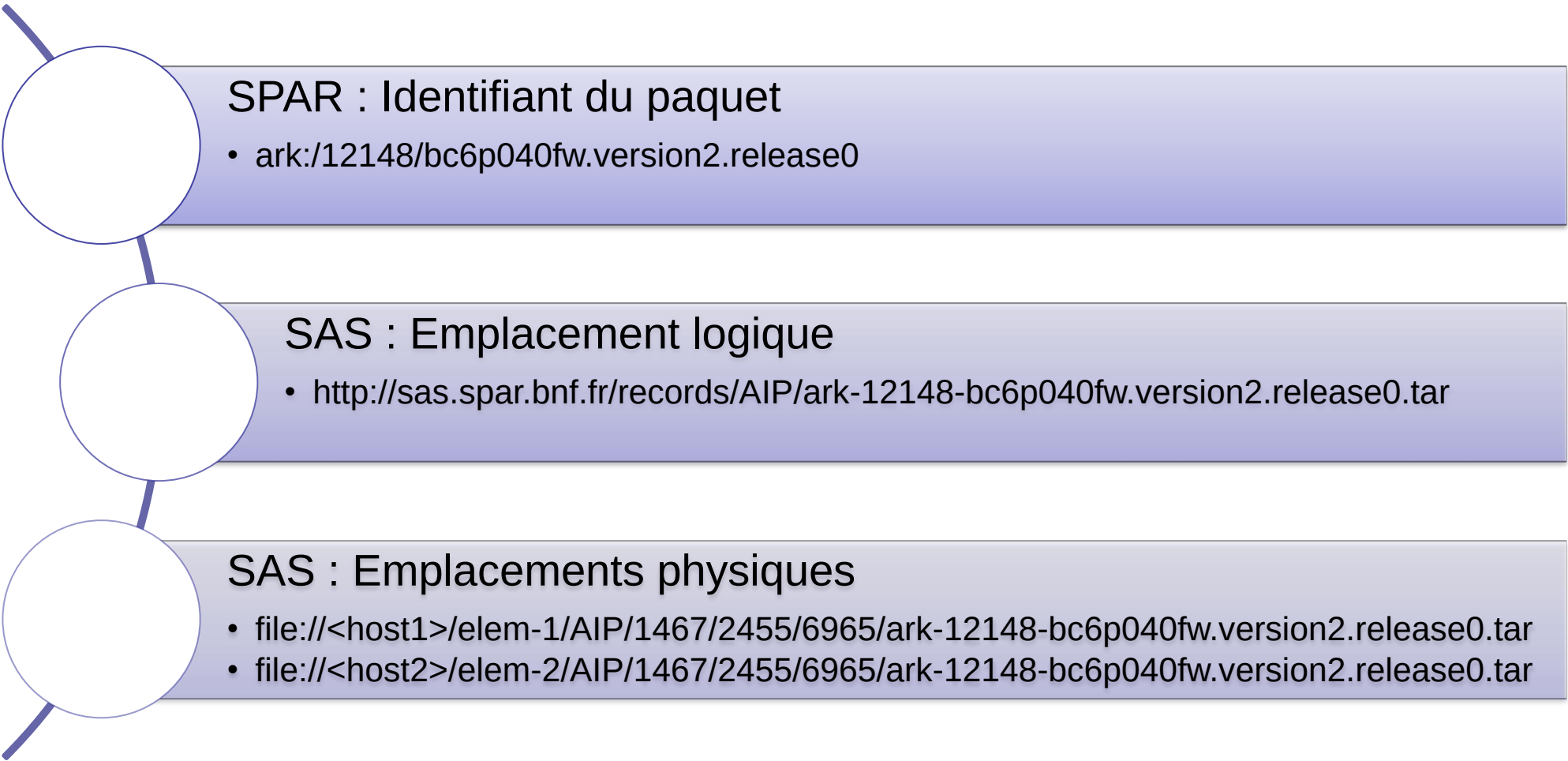
Administration > Surveillance matérielle > Suivre les capsules

Suivre les capsules ?			
Capsule	Statut	% occupation	Nb enregistrements
aip-ref	ONLINE		93
capsAIP01	READONLY		149463
capsAIP02	OFFLINE		2783
capsAIP03	OFFLINE		717
capsAIP04	ONLINE		31041
capsAIP05	OFFLINE		1

Implémentation technique : iRODS



Identification et localisation



SPAR : Identifiant du paquet

- `ark:/12148/bc6p040fw.version2.release0`

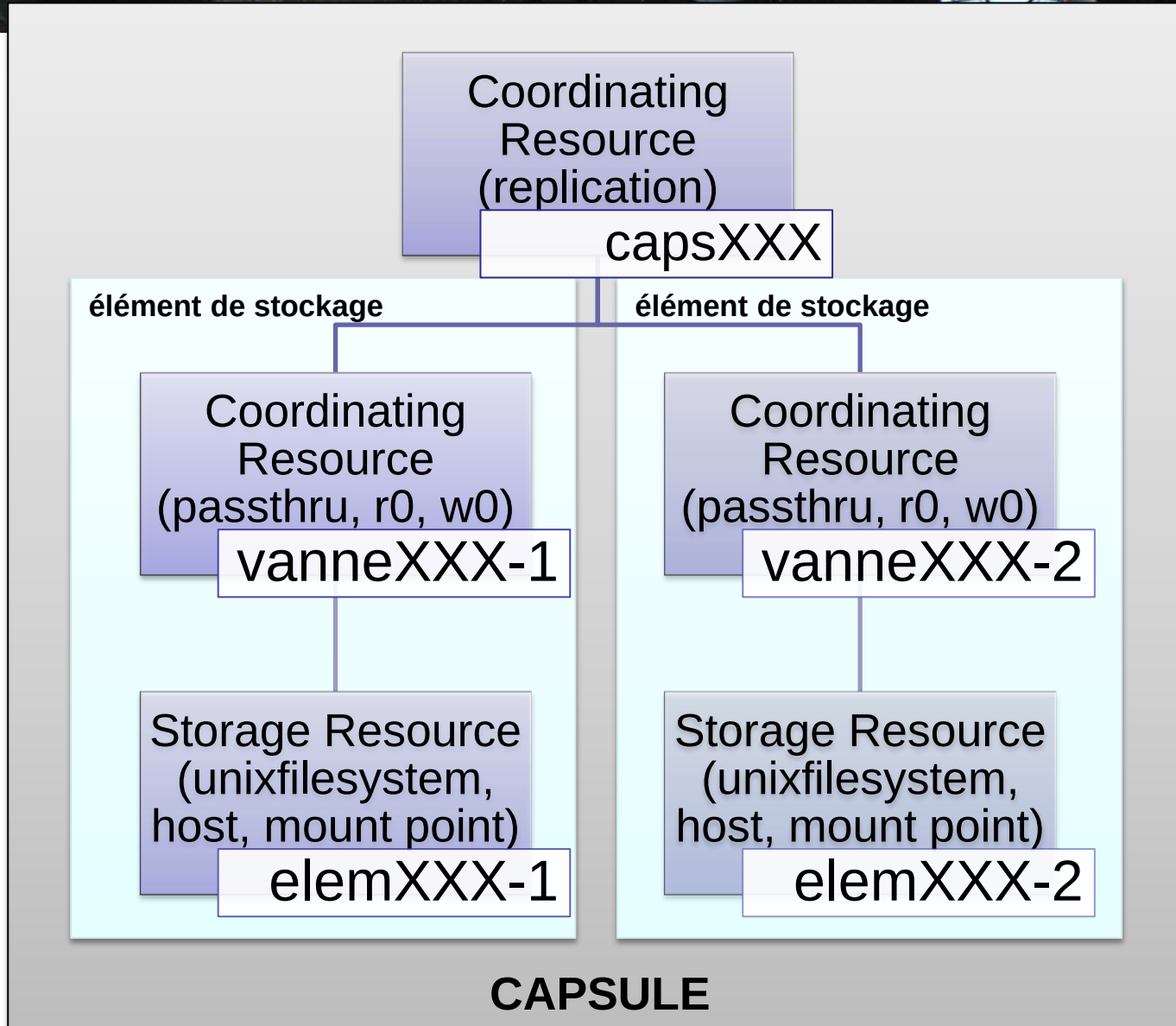
SAS : Emplacement logique

- `http://sas.spar.bnf.fr/records/AIP/ark-12148-bc6p040fw.version2.release0.tar`

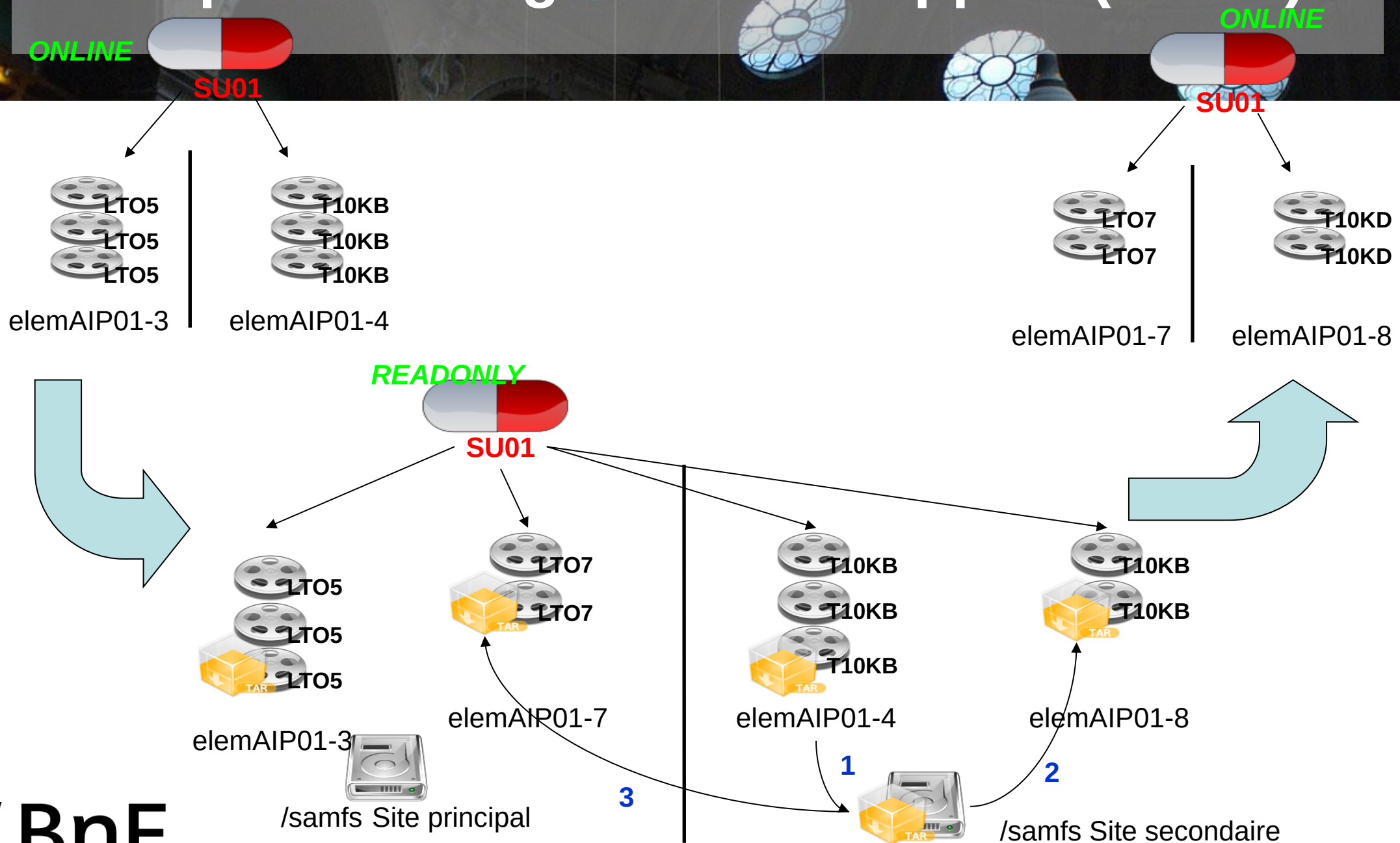
SAS : Emplacements physiques

- `file://<host1>/elem-1/AIP/1467/2455/6965/ark-12148-bc6p040fw.version2.release0.tar`
- `file://<host2>/elem-2/AIP/1467/2455/6965/ark-12148-bc6p040fw.version2.release0.tar`

SAS ↔ iRods — Capsules et éléments de stockage



Principe de la migration de support (SPAR)



Migrations réalisées

- 2014 : 9840C (40GB) -> LTO5 (1.5TB), T10000A (500GB) -> T10000B (1TB), opération de 3 mois
- 2017-2018 : LTO5 -> LTO7 (6TB), T10000B->T10000D (8.5TB), opération de 1,5 an pour 1,5Po, taux de réduction de bandes de 5
- Leçons :
 - opérations longues qui nécessitent un suivi continu,
 - suivi « à la Excel » pas efficace

Infrastructure de stockage au 30/10/2018

Site 1

2 x SL8500



3,5Po

Bandes utilisées



LTO6 593

LTO7 358



Site 2

2 x SL8500



3,5Po

Bandes utilisées



T10000C 348

T10000D 314



Veille sur l'obsolescence technologique

- Bandothèque SL8500
 - Originellement, conçue par StorageTek (société fondée en 1969)
 - Rachat par Sun Microsystems en 2005 (4,1MM\$)
 - Oracle Corporation rachète SUN le 20 avril 2009 (7,4MM\$)

Drive	T10000A	T10000B	T10000C	T10000D	T10000E
Year	2006	2008	2011	2013	Never
Capacity	500GB	1TB	5TB	8TB*	N/A
Performance	120 MB/s	120 MB/s	240 MB/s	252 MB/s	N/A

- 2019 : décision BnF d'acquérir un nouveau matériel

Infrastructure : bandothèques IBM





Mutualisation des infrastructures

Archive SA (Système Audiovisuel)



Robotique GRAU
en fin de vie (2021)

1,2 millions de boîtes

2,5 Po de données

1 219 cartouches LTO-6

Infrastructures existantes en 2020



bandes LTO6



bandes LTO4



Bandes LTO6

GRAU



bandes LTO6 et 7



bandes T10000C et D

ORACLE



bandes LTO8



bandes Jaguar

IBM

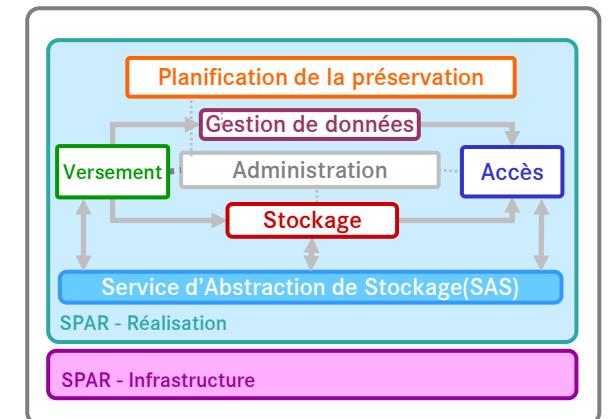
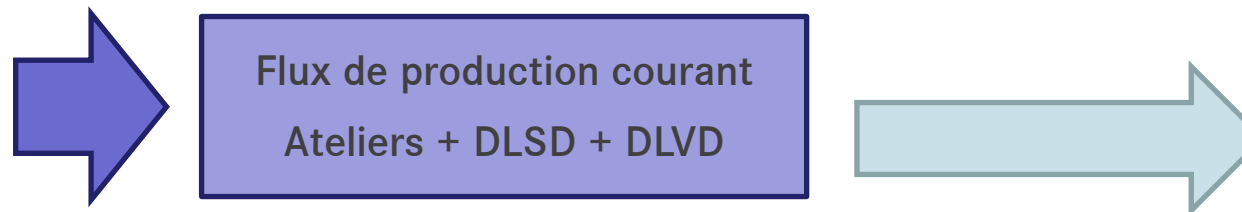
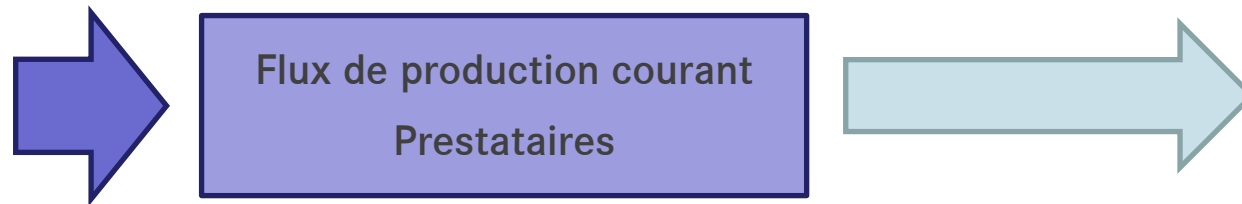
SITE 1

SITE 2

Etape 1 : Flux courant vers l'IBM

Il s'agit de faire en sorte que les flux d'entrée courant aillent sur la bandothèque IBM :

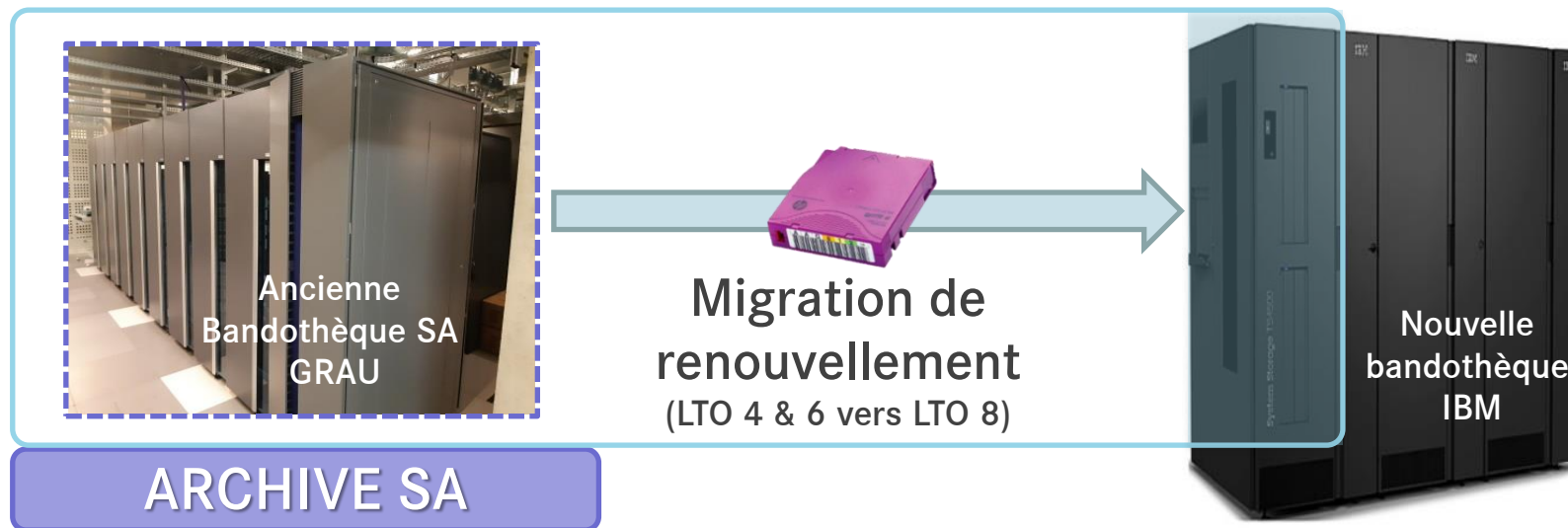
- 1/ Flux SA => copie directe des boites sur la bandothèque
- 2/ Flux DL dématérialisé => stockage dans SPAR



Etape 2 : Renouvellement du GRAU vers IBM

Il s'agit de recopier l'ensemble des données stockés sur le GRAU vers l'IBM

Durée prévue : 2 ans !!!



Réalisé en 7 mois grâce à ...

{ BnF

SA Deck : Outil de suivi du renouvellement de la bannothèque GRAU vers l'IBM

{SA-DECK (Administrateur)

Flower API : KO

Version : 1.1.0PFC50

Suivi

Bilans

Estimations / Priorisations

Administration

Informations

Campagne de renouvellement : Renouvellement des bandes du SA

Titre : Renouvellement des bandes du SA

Début : 29/04/2021

Etat de la campagne : En cours

Description : Copie des bandes LTO6 (SL8500) vers LTO8 (TS4500)

En attente 1

En cours

Terminé

Boites du SA 1

Capsule

Etat de migration

Enregistré

capsule

En attente

Démarrer

1 247

Statistiques

Etat des capsules

En attente (1)

Détail des éléments de stockage

Nom	Site	Chemin	Bande	Bannothèque	Avancement (taille)	Avancement (enregistrements)
elemSA-4	Tolbiac	Tolbiac /dev/stn[x]	LTO4	GRAU Storage	0 octets / 2,5 Po 0 %	0 / 1 245 071 0 %
elemSA-6	Tolbiac	Tolbiac /dev/stn[x]	LTO6	Oracle SL8500	0 octets / 2,5 Po 0 %	0 / 1 247 336 0 %

Détail des bandes de elemSA-6 1219

Filtres

Etat de la bande :

Nom de la bande :

Filter

1 219 bandes trouvées, affichage de 1 à 50. [Premier/Précédent] 1, 2, 3, 4, 5, 6, 7, 8 [Suivant/Dernier]

Nom	Etat	Priorité	Début	Fin	Durée	Avancement (taille)	Avancement (enregistrements)
Z00001	En attente					0 octets / 2,1 To 0 %	0 / 1 076 0 %
Z00002	En attente					0 octets / 2,1 To 0 %	0 / 938 0 %
Z00003	En attente					0 octets / 2,1 To 0 %	0 / 886 0 %
Z00004	En attente					0 octets / 2,1 To 0 %	0 / 1 163 0 %
Z00005	En attente					0 octets / 2,1 To 0 %	0 / 1 639 0 %
Z00006	En attente					0 octets / 2,1 To 0 %	0 / 1 106 0 %

SA Deck : Opération de suivi

SA-DECK (Administrateur)

Version : 1.1.0

Suivi

Bilans

Estimations / Priorisations

Administration

Informations

Capsule CAPSULESA

1 En attente

2 En cours

3 Terminé

Nom : capsuleSA

Etat de migration :

En cours

Nombre d'enregistrements enregistré : 1 228 976

Taille enregistrée : 2,5 Po (2 472 130 473 866 382 octets)

Avancement (Enregistrements) : 42 / 1 228 976

Avancement (Taille) : 165,9 Go / 2,5 Po

Campagne : Renouvellement des bandes du SA (22/06/2021) : Copie des bandes LTO6 (SL8500) vers LTO8 (TS4500)

Espace réservé au 'staging' : 0 octets (0 octets)

Espace de 'staging' utilisé : 0 octets (0 octets)

Début migration : 23/06/2021 17:07:19

Durée : 02:48:23

Nombre de lecteurs max : 5

Fin migration :

Temps restant estimé : 11 jours 13:50:22

Modifier la configuration

Pause

Détail des éléments de stockage

Nom	Site	Chemin	Bande	Bandothèque	Avancement (taille)	Avancement (enregistrements)
elemSA-4	Tolbiac	Tolbiac /dev/stn[x]	LTO4	GRAU Storage	0 octets / 2,5 Po 0 %	0 / 1 226 711 0 %

Démarrage le 23 juin 2021. Fin le 1^{er} février 2022.

{ BnF

Mise en œuvre de l'outil SADeck

Capsule CAPSULESA

1 En attente

2 En cours

3 Terminé

Nom : capsuleSA

Etat de migration : En cours

Nombre d'enregistrements enregistré : 1 228 977

Taille enregistrée : 2,5 Po (2 472 136 582 469 553 octets)

Avancement (Enregistrements) : 1 169 592 / 1 228 971

(Enregistrements) :

95 %

Avancement (Taille) : 2,3 Po / 2,5 Po

94 %

Campagne : Renouvellement des bandes du SA (22/06/2021) : Copie des bandes LTO6 (SL8500) vers LTO8 (TS4500)

Espace réservé au 'staging' : 0 octets (0 octets)

Nombre de lecteurs max : 9

Espace de 'staging' utilisé : 0 octets (0 octets)

Début migration : 23/06/2021 17:07:19

Fin migration :

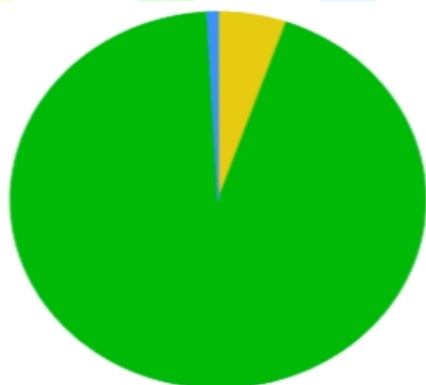
Durée : 211 jours 16:30:02

Temps restant estimé : 13 jours 12:17:14

Statistiques

Etat des bandes pour la capsule

En attente (63) Terminé (1111) En cours (10)



Etat des enregistrements pour la capsule

En attente (59270) En cours (10) Terminé (1189694) Erreur (1)



Etape 3 : MIGRATION DES CONTENUS SA VERS SPAR

Il s'agit de verser dans le Système de Préservation et d'Archivage Réparti les données de l'archive SA.

Migration réalisée par type de contenu

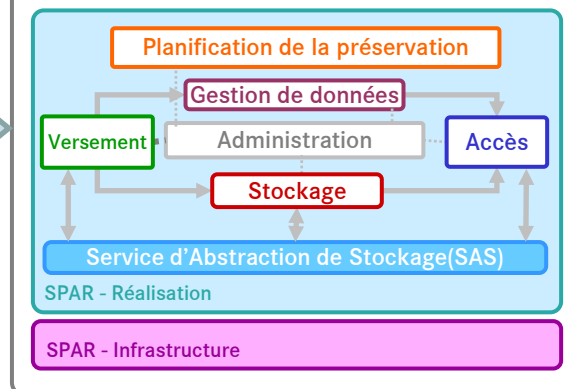
Migration de
ré-empaquetage

3

Rétrospectif AUD
(SDC 12, VK, SM,
VDVD,...)



SPAR



Création de l'outil SelectSABox

Select SA Box

+ Importer

Boîtes Exemplaires Numériques Exemplaires Originaux Lots Rapport

Build Infos

Liste des exemplaires numériques

Appuyer pour lancer le traitement de complétion :

▶ Commencer

1 057 593 resultat(s) sur 42 304 page(s)

	Exemplaire Numérique				Préfixe	Reste à faire
Exemplaire Numérique	Nb. Volumes	Nom de Variante	Cote	ARK Notice	Préfixe	Nb. Logs
1080372	12	K7A			8MU_000027	
1080373	9				8MU_000027	10
1080374	18	K7A			8MU_000028	10
1080375	14				8MU_000028	19
1080376	20	K7A			8MU_000030	15
1080377	17				8MU_000030	21
1080378	17	K7A			8MU_000031	18

Objectif : donner une destination à chaque boîte SA

- L'outil **SelectSABox** permet de définir des **lots homogènes de traitement** pour les basculer dans une filière de traitement.
- Pour chaque lot, on détermine :
 - son signalement (catalogue général ou BAM)
 - son mode d'entrée et sa filière dans SPAR (DL, ADDN, Numérisation...)
 - le réempaquetage à opérer pour transformer la boîte SA en paquet d'archive
 - son mode de diffusion : PAV RdJ, Gallica ou GIM

ARCHIVE NUMERIQUE AUDIOVISUELLE -



SDC12_095342_V1_2 Intégrale



SDC12_095342_V2_2 Intégrale

Lettrage : SDC12 + 6 digit ✓

Variante : sans ✓

Volumaison : Vn_m ✓

Types fichiers : ✓

SDC12_095342_V1_2_R.tif

SDC12_095342_V1_2_V.tif

SDC12_095342_V1_2_01.wav

SDC12_095342_V1_2_02.wav

SDC12_095342_V1_2_03.wav

SDC12_095342_V1_2_04.wav

SDC12_095342_V1_2_05.wav

SDC12_095342_V1_2_06.wav

SDC12_095342_V1_2_07.wav

SDC12_095342_V1_2_08.wav

Exemple : SDC12-95342



{BnF Catalogue général

Tout

Recherche avancée

AUTEURS A-Z

SUJETS A-Z

Accueil > Recherche avancée > Liste de notices > Notice bibliographique

Notice bibliographique

Notice

Au format public

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

Interprète(s) : [Dubas, Marie \(1894-1972\)](#). Chant

Titre(s) : Intégrale [Enregistrement sonore] : 1927-1945 / Marie Dubas, chant... [et al.]

Publication : Vincennes : Frémeaux & Associés ; La Plaine St Denis : distrib. Night & Day, 1996 (P)

Éditeur : [Frémeaux & Associés](#). Éditeur commercial

Distributeur : [Night & Day](#). Distributeur

Producteur(s) : [Frémeaux & Associés](#). Producteur de phonogrammes

Description matérielle : 2 disques compacts (1 h 13 min 59 s, 1 h 14 min) + 1 brochure ([39] p.) : ill. ; 12 cm ✓

Note(s) : Repiquages et références commerciales des enregistrements d'origine. - Vincennes : prod. Frémeaux & Associés, P 1996

Notices / Pénet, Martin ; Bellair, François

Notice : [ark:/12148/cb38355349d](https://nla.ark:/12148/cb38355349d) ✓

Filière SPAR : Fil_AUD_B ✓

LOCALISER CE DOCUMENT
(2 EXEMPLAIRES)

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

NUMAV-48777 ✓

support : son dématérialisé

Réserver

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

SDC 12-95342 ✓

support : CD-audio 12 cm

Voir détail exemplaire

indisponible : document non diffusable

PEX de l'exemplaire original :
SDC 12-95342 ✓

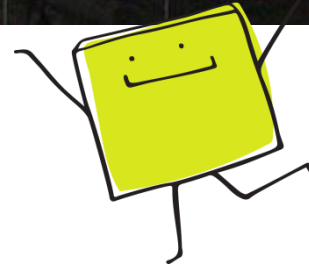
PEX de l'exemplaire numérique :
NUMAV-48777 ✓

Lot x Validé ✓

OK pour ingestion dans SPAR

Phasage des lots par lettrage de cote

SDC12	
Chantier bib	Réempaquetage



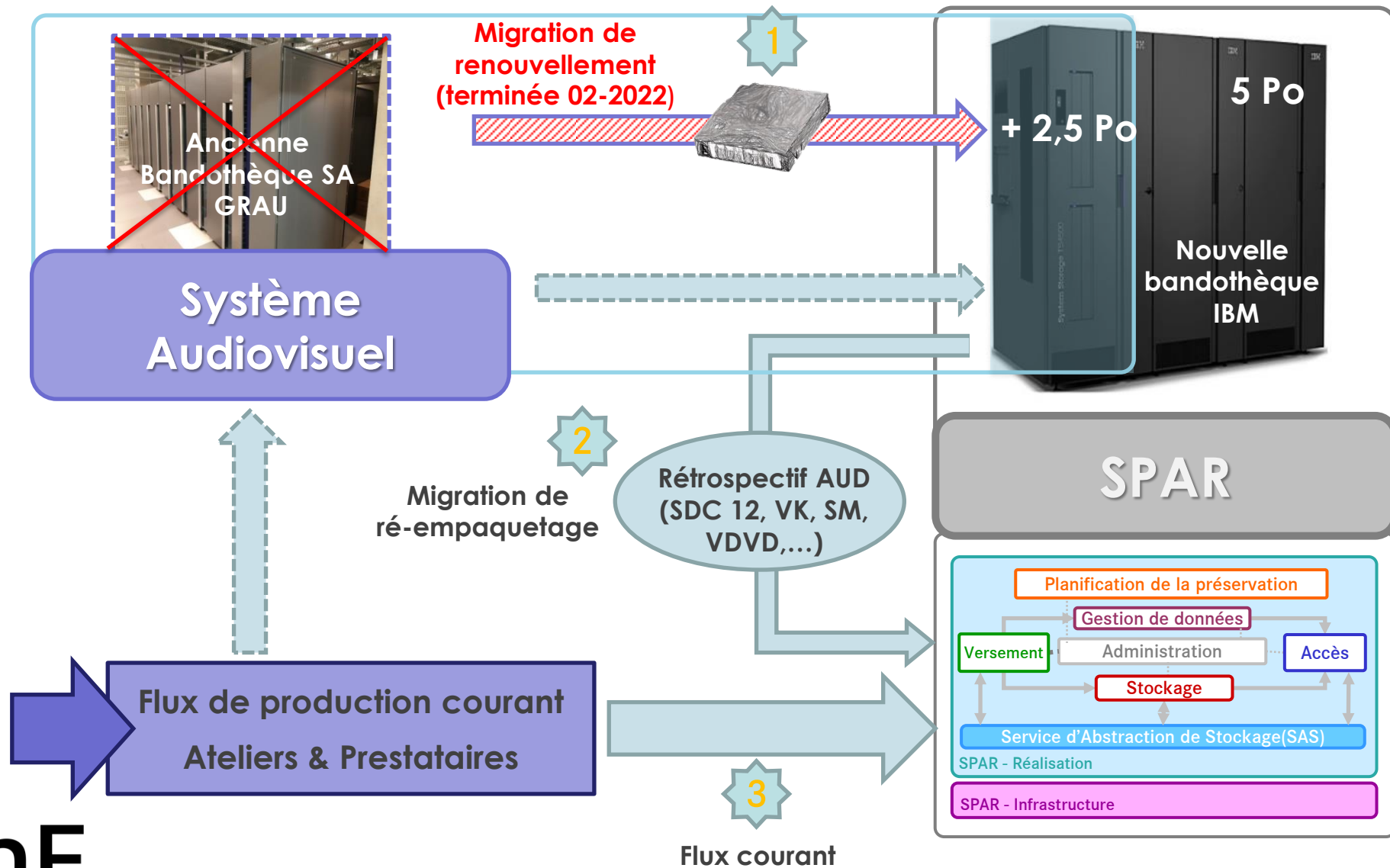
VK	
Chantier bib	Réempaquetage

VDVD	
Chantier bib	Réempaquetage

SM	
Chantier bibliographique	Réempaquetage

{ BnF

Trajectoire globale de la migration





Merci de votre attention

thomas.ledoux_AT_bnf.fr