



La plateforme PAC et ses évolutions fonctionnelles et techniques

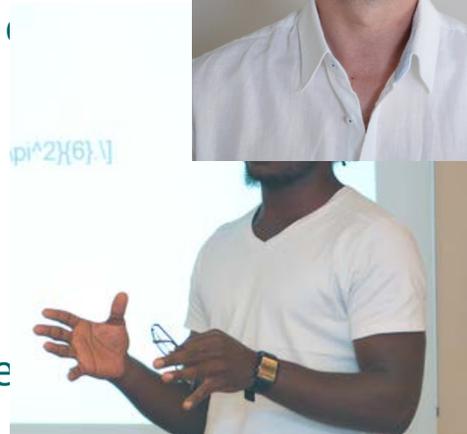
3^{ème} journée des utilisateurs de l'archivage – 9 juin 2015



- **L'équipe d'archivage du CINES**
- **Les évolutions dans PAC entre juin 2014 et juin 2015 :**
 - **Renforcement de la sécurité**
 - **Améliorations sur la validation des formats de fichiers**
 - **Nouvelle interface de FACILE**
 - **Autres évènements à signaler**
- **Les évolutions à venir :**
 - **Evolution de l'architecture de stockage**
 - **Nouveaux formats d'archivage : TEI archivable et formats 3D**
 - **Projets d'archives en 2015/2016**
- **Annonces diverses**

- ✓ **Marion Massol** : Responsable du Département Archivage et Diffusion (DAD)
- ✓ **Lorène Béchar** : archiviste et responsable fonctionnel PAC
- ✓ **Jean-Pierre Théron** : administrateur plateforme et responsable technique PAC
- ✓ **Mireille Gay** : administrateur outils internes et référent projet
- ✓ **Benjamin Watine** (Logixys) : développeur, consultant
- ✓ **Maurice Gaujac** : assistance FACILE 2, gestionnaire statistique projet
- ✓ **Alexandre Granier** : développeur, expertise formats
- ✓ **Franklin Boumda** : développeur, expertise formats
- ✓ **Pascal Dugénie** (Camanoë) : ingénieur projets
- ✓ **Michel Bouloc** : référent projet

- ✓ **De nouveaux arrivants :**
 - Alexia de Casanove (Logixys) : archiviste
 - Alexis Jean-Laurent : responsable données scientifiques projet
 - Adrien Quilleret (Camanoë) : développeur



- ✓ **Mireille Gay** : ABES, INSERM, BUPMC, BIUS
- ✓ **Jean-Pierre Théron** : PERSEE, BSG, IRSTEA, GEOSUD
- ✓ **Michel Bouloc** : Cour des Comptes, CUJAS, INRAP
- ✓ **Maurice Gaujac** : Université de Lorraine
- ✓ **Alexis Jean-Laurent** : SCD Lille 2
- ✓ **Alexandre Granier** : Sciences Po, INRA
- ✓ **Nicolas Larrousse (TGIR Huma-Num)** : CCSD, EFEO, Cocoon, Archéovision, Cléo, ATILF, MMSH, Ortolang, IRHT

et **Alexia de Casanove** sur l'ensemble des projets pour les questions archivistiques



Les évolutions depuis 1 an

Le renforcement de la sûreté et de la sécurité



- **Amélioration de l'identification :**
 - **Mise à jour signatures Droid v81**
 - **Formats basés sur le XML**
- **Nouveaux formats :**
 - **TEI Text Encoding Initiative (anciennement validable en XML)**
- **Changement de validateurs (exit mencoder)**
- **Prochaines modifications**
 - **Mise en place de MkValidator**

- Nouveau Design CINES Compliant
- Envoi de plusieurs fichiers simultanément
- Aide à la validation pour certains formats
- Indication « Archivable sur PAC »
- Documentation Web Service
- Perspectives :

- Nouveaux tutoriels
- Conversion de format
- Validation de paquets PAC

FACILE - Service de validation de formats
Vérifier l'éligibilité de vos documents à un archivage sur la plateforme PAC du CINES.

+ Choisissez des fichiers - Taille max 2,5 Go Valider Annuler

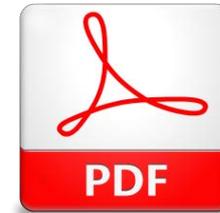
Cliquez ici pour demander l'aide d'un expert du CINES

Liste des formats validables

⚠ Attention : le validateur de formats permet de valider certains formats qui ne sont pas pris en charge par la plateforme d'archivage du CINES.

Format	Nom	PRONOM PIUD	Type MIME	Commentaire	Archivable dans PAC
AAC AAC	Advanced Audio Codings	[fmt/199]		format MPEG-4 contenant uniquement un flux audio au format AAC	✓
AIFF PCM	Audio Interchange File Format	[fmt/414]	[audio/x-aif, audio/x-aiff]	format audio contenant uniquement un flux PCM	✓
FLAC FLAC 1.2.1	Free Lossless Audio Codec	[fmt/279]	[audio/ogg, audio/x-flac]	format audio compressé sans perte	✓
GIF 87a	Graphics Interchange Format	[fmt/5]	[image/gif]	format image pouvant contenir également des animations	✓
GIF 89a	Graphics Interchange Format	[fmt/4]	[image/gif]	format image pouvant contenir également des animations	✓
GeoTIFF	Geographic Tagged Image File Format	[fmt/155]	[image/tiff]	format TIFF contenant des informations de géoréférencement et de géolocalisation	✓
HDPS 1.0	Hierarchical Data Format	[fmt/286]		format de données à caractère scientifique	✗
HDPS 2.0	Hierarchical Data Format	[fmt/287]		format de données à caractère scientifique	✗
JPEG	Joint Photographic Experts Group	[fmt/41, fmt/44]	[image/jpeg]	format de représentation compressée d'une image fixe.	✓
JPEG2000	JPEG 2000	[fmt/151, x-fmt/392]		extension du format JPEG standard	✓
JPEG 1.00	Joint Photographic Experts Group	[fmt/42]	[image/jpeg]	format de représentation compressée d'une image fixe.	✓
JPEG 1.01	Joint Photographic Experts Group	[fmt/43]	[image/jpeg]	format de représentation compressée d'une image fixe.	✓

- ✓ Publication de la partie 3 de l'étude PDF sur le test des outils de validation du PDF



- ✓ Renouvellement du DSA (Data Seal of Approval)





Les évolutions à venir

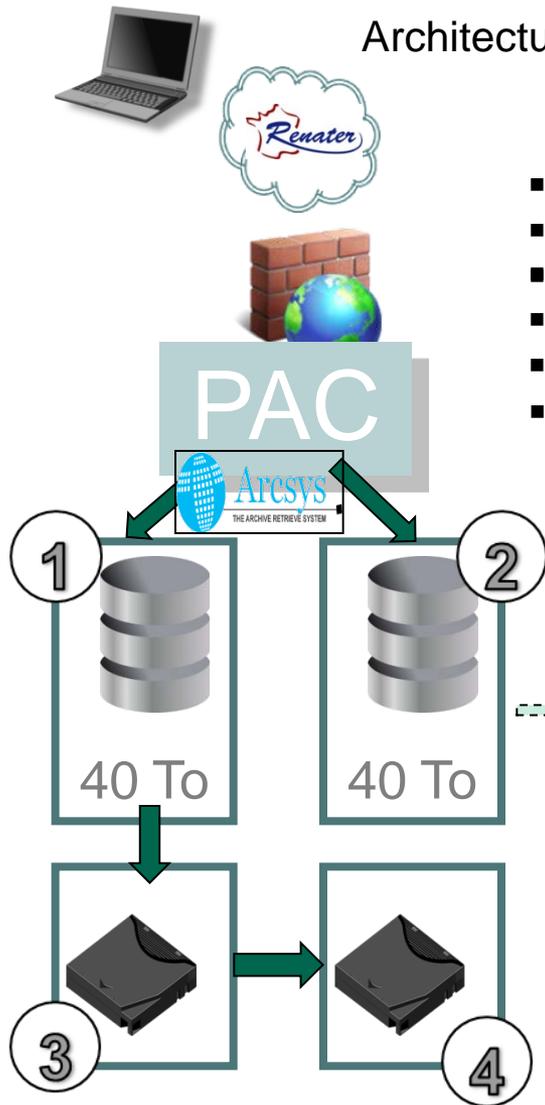
- ✓ La préservation pérenne doit prendre en compte la dégradation des supports physiques et des outils d'exploitation

- ✓ Après 7 ans d'exploitation, l'infrastructure en place doit être remplacée :
 - Remplacement des supports physiques : Baie de disques
 - Remplacement des serveurs du SAE : vers une architecture virtualisée

- ✓ Des évolutions fonctionnelles du SAE : diversification des offres du service d'archivage

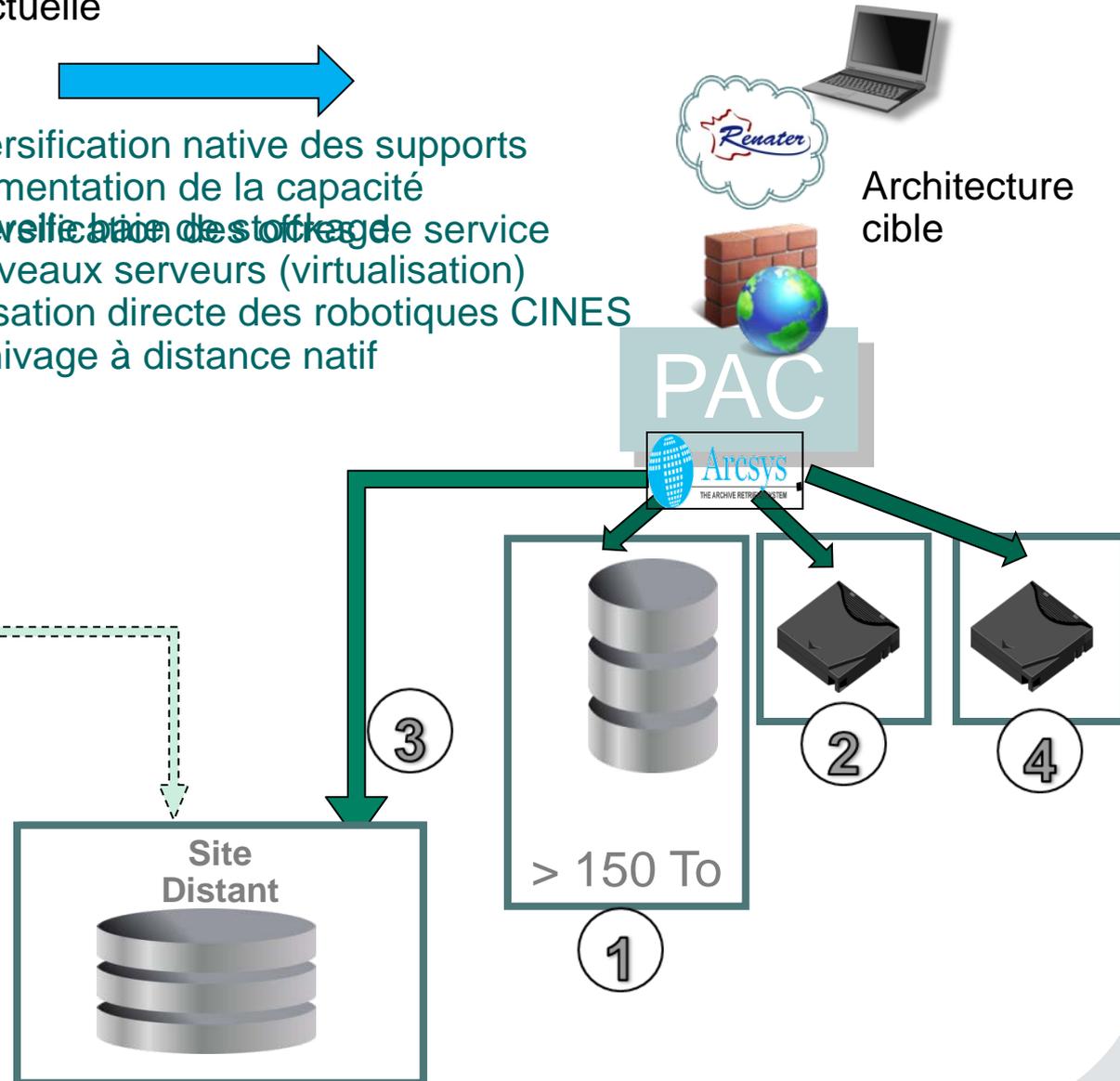
Evolution de l'architecture de stockage

Architecture actuelle



- Diversification native des supports
- Augmentation de la capacité
- Diversification de stockage de service
- Nouveaux serveurs (virtualisation)
- Utilisation directe des robotiques CINES
- Archivage à distance natif

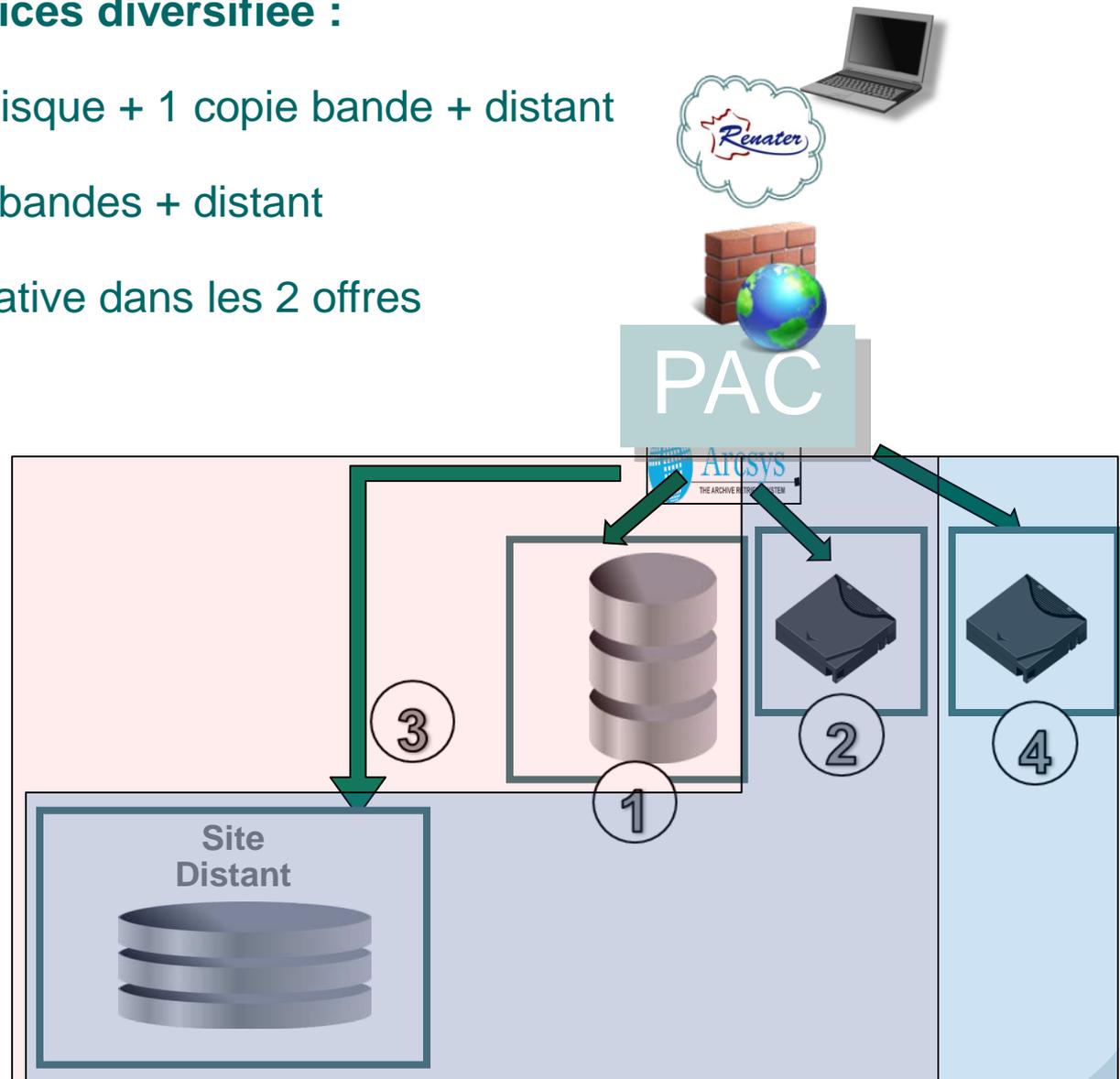
Architecture cible



Evolution de l'architecture de stockage

✓ Vers une offre de services diversifiée :

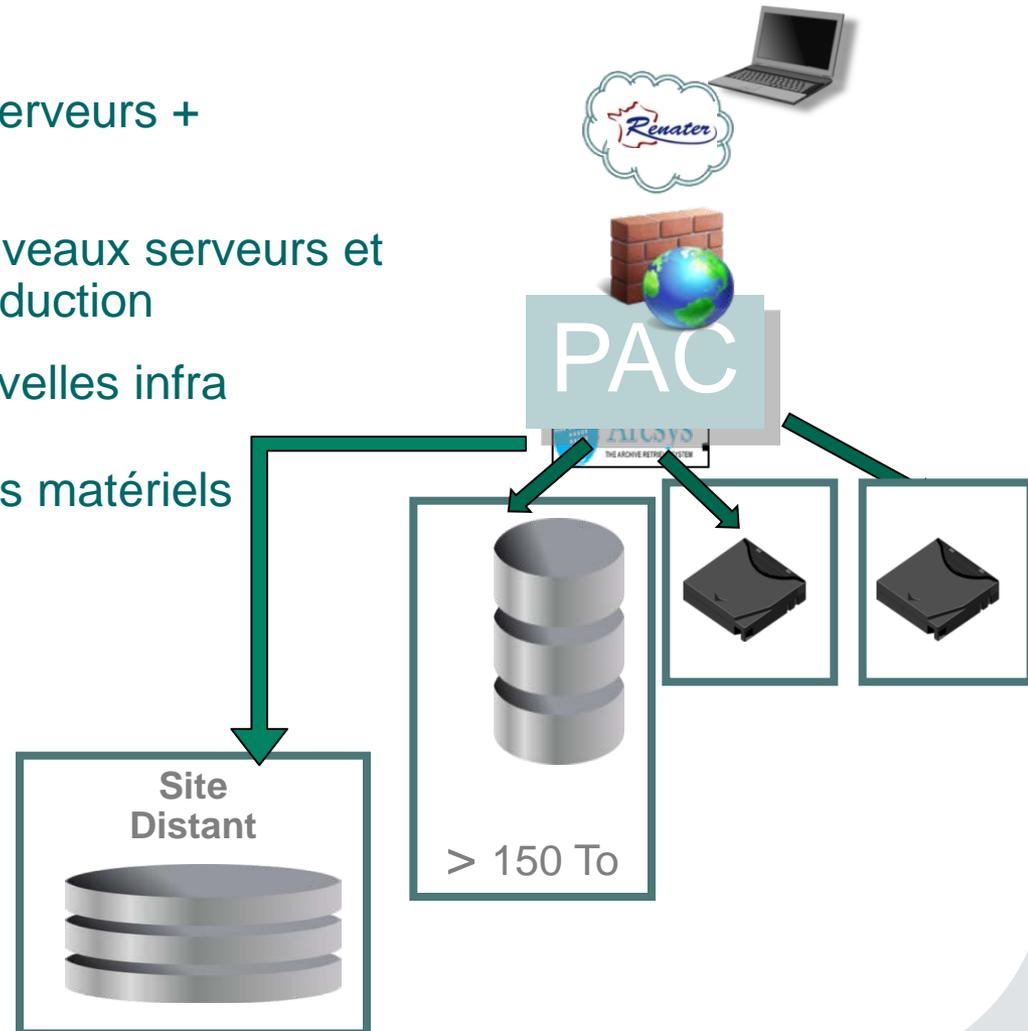
- Archivage : 1 copie disque + 1 copie bande + distant
- Archivage : 2 copies bandes + distant
- Une copie distante native dans les 2 offres



Evolution de l'architecture de stockage

✓ Mise en œuvre de la migration :

- Travaux de développement de l'interface entre SAE et Robots CINES
 - Acquisition des matériels : Baie + Serveurs + Lecteurs de bandes (en cours)
 - Installation / Configuration des nouveaux serveurs et de la nouvelle baie → arrêt de production
 - Migration des données sur les nouvelles infra
 - Dé-commissionnement des anciens matériels
- Objectif : Fin 2015



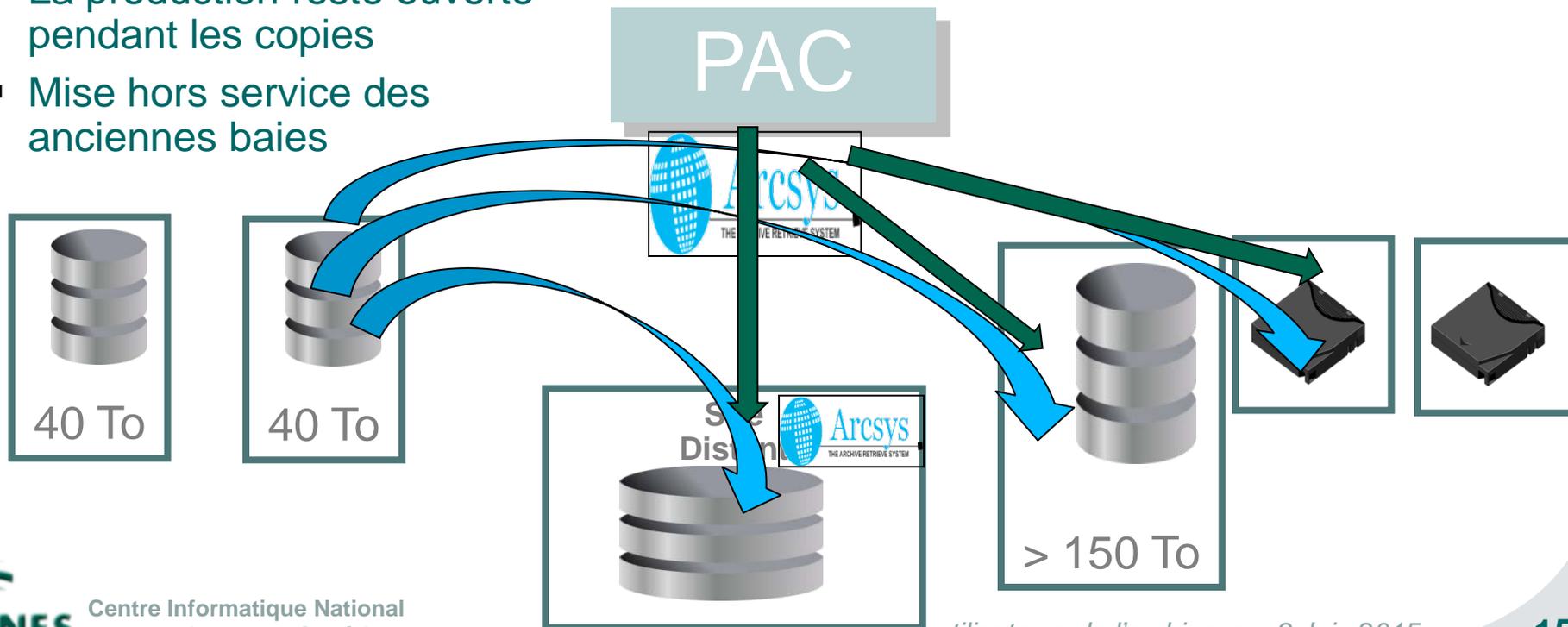
Evolution de l'architecture de stockage

✓ Mise en œuvre de la migration :

- 3 flux de copies
- Gérés par le SAE
- Conservation de la traçabilité
- Contrôles d'intégrité
- Base d'archives / Base d'archives
- La production reste ouverte pendant les copies
- Mise hors service des anciennes baies

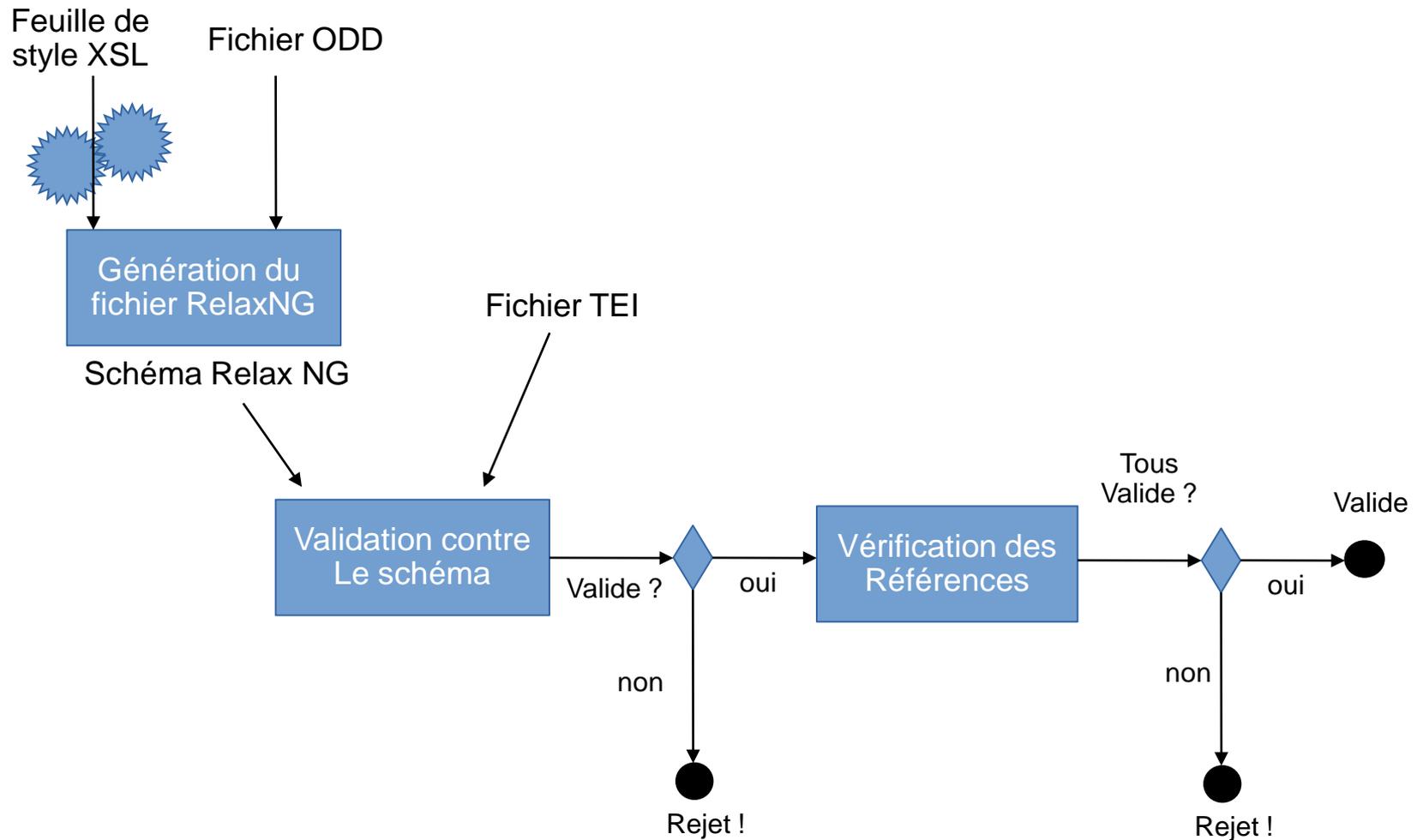


Architecture intermédiaire de migration



- Le format TEI sert à coder sous forme numérique toutes sortes de textes
- Basé sur le format XML
- Une grammaire spécifique définie dans le schéma `tei_all.xsd` permet de réaliser un fichier ODD qui contient la documentation et permet de générer un schéma pour valider le corpus
- Le validateur de format permet de valider un fichier TEI contre son fichier ODD (One Document Does it all)

TEI : principe de validation



- Suite à une demande d'Archeovision pour la numérisation de sites archéologiques en 3D
- Deux formats sélectionnés :
 - Collada (basé sur XML)
 - PLY Polygon File Format
- Disponibles dans la prochaine version du validateur (automne 2015)

✓ Projets d'archivage en cours :

IRSTEA : pilote sur les publications

GEOSUD : images satellites

PRES Université de Lorraine : thèses numérisées

INRAP : rapports de fouilles archéologiques

INSERM-ANRS : dossiers patients des études cliniques

SCD Lille 2 : thèses et mémoires

Sciences Po : fonds numérisés (monographie, archives électorales, dossiers de presse)

Archéovision : données archéologiques en 3D



Institut national
de recherches
archéologiques
préventives



✓ Projets d'archivage à venir :

ATILF : ouvrages encodés en TEI

CLEO : contenu de Open Edition Books

CNRS : dossiers de carrière des agents

INRA : données de recherche de phénotypage végétal

IRHT : manuscrits numérisés

MMSH : phonothèque

Ortolang : recherches linguistiques en TEI



- ✓ Nouveau site Web du CINES depuis l'été 2014
- ✓ L'actualité du SEDA 2.0
- ✓ Les déménagements de matériels à venir

The screenshot shows the homepage of the CINES website. At the top, there is a navigation bar with links for 'Marchés publics', 'Historique', 'FAQ', and 'Français', along with a search box. Below this is the CINES logo and a banner with three images of the facility. A secondary navigation menu includes 'Accueil', 'Présentation', 'Calcul', 'Archivage', 'Services', 'Formations', 'Actualités', 'Annonces', and 'Contact'. A 'FLASH INFO' section highlights 'PDF/A l'essentiel 2.0, la norme, les out_'. The main content area features a large image of server racks with the title 'Archivage pérenne' and a brief description. To the right, a sidebar lists various services like 'Recrutement et stages', 'Documentation technique', and 'Ressources'. Below the main article, there are several news items with expandable sections: 'Le CINES', 'Le CINES & L'Europe', 'Comité Des Utilisateurs', 'OCCIGEN : Le Nouveau Supercalculateur', 'Formation PATC PA04', 'EUDAT une solution simple pour la recherche des données', 'Renouvellement du Data Seal of Approval (DSA)', and 'Formation CIO2'.



Questions & Réponses

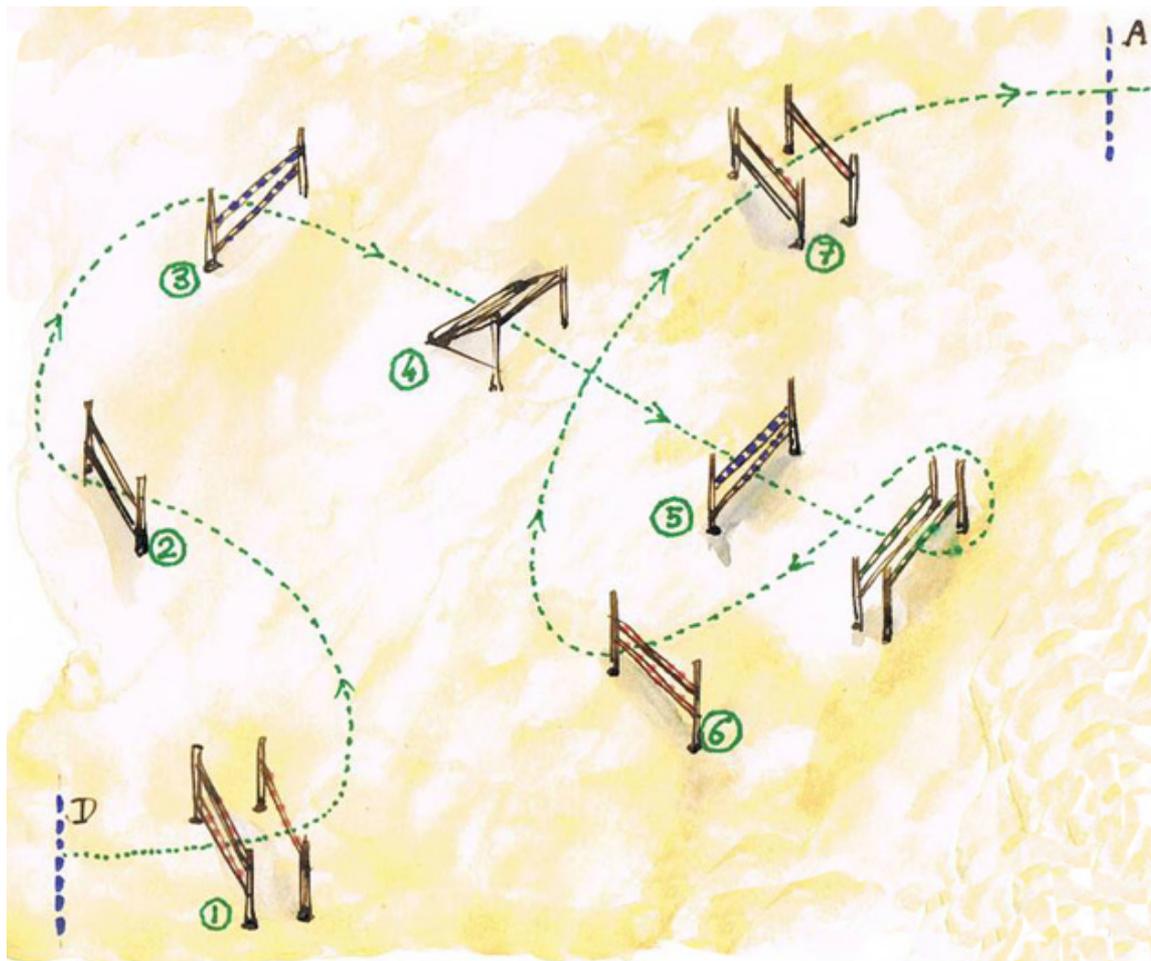


Echanges autour des problématiques rencontrées pendant la phase d'avant-projet

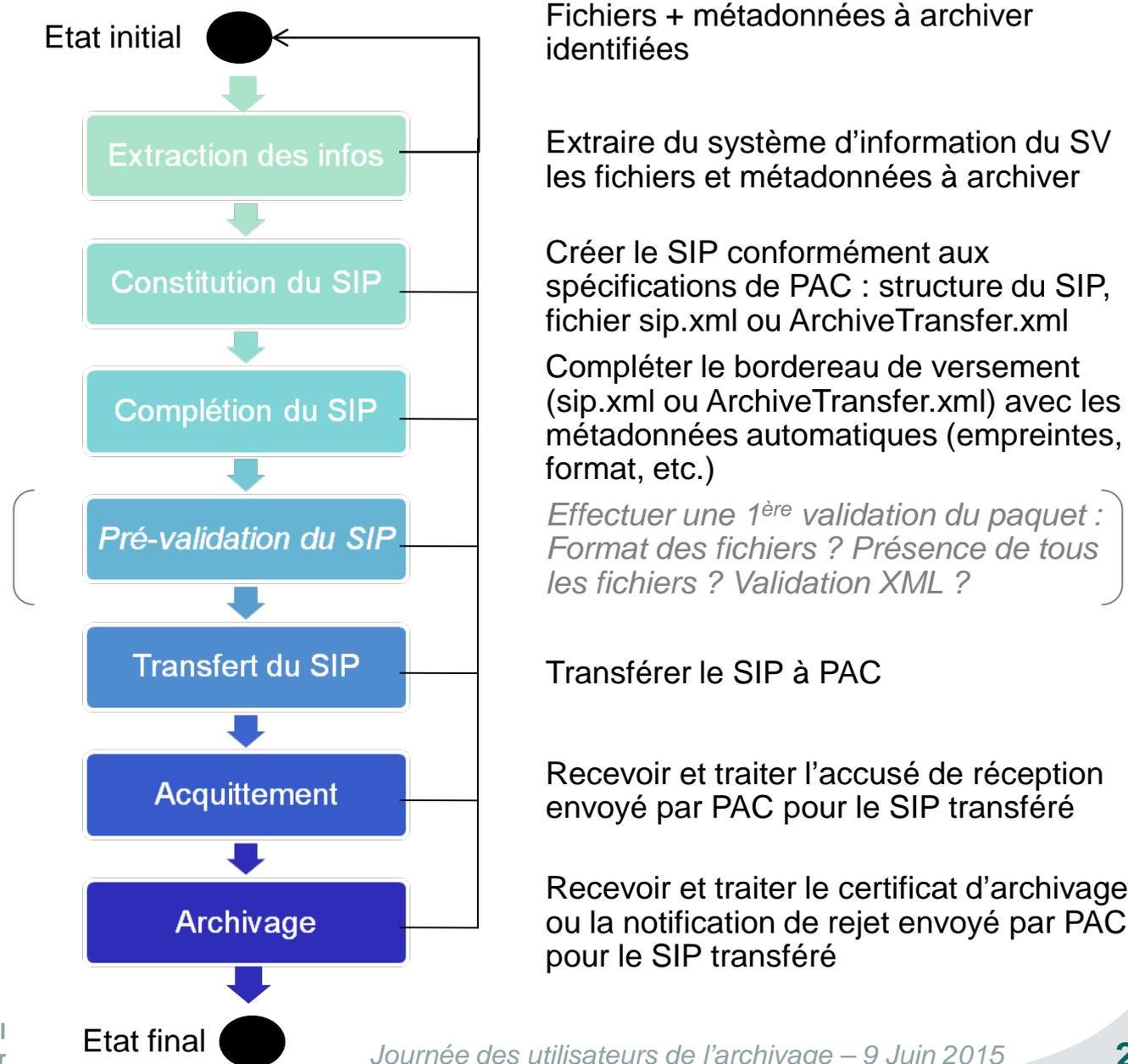
3^{ème} journée des utilisateurs de l'archivage – 9 juin 2015



Quels ont été / sont les points d'achoppement lors de la mise en œuvre des projets d'archivage au CINES ?



Le processus de versement au CINES côté Service Versant





Echanges sur les évolutions possibles de l'offre d'archivage du CINES et questions diverses

3^{ème} journée des utilisateurs de l'archivage – 9 juin 2015

