



DASD - Département Archivage & Services aux Données

Archivage retour d'expérience et étude des coûts

R.I.P Data

02 Octobre 2024 @ Philippe PRAT

Le CINES en quelques mots

Basé à Montpellier

- 60 personnes : ingénieurs, techniciens, archivistes, administratifs;
- 45 ans d'existence - *CNUSC en 1980 puis CINES en 1999*;
- Sous la tutelle du Ministère de l'Enseignement supérieur et de la Recherche (MESR).

Fournit à la communauté ESR des 3 missions nationales pour l'ESR:

- Calcul intensif : adastr (1+2) 90 Pflops, scalaire et convergé pour l'IA et le passage à l'exascale;
- 20 ans d'expérience dans l'archivage (26 projets en production , 14 en étude, 100 To archivé) ;
- Hébergement.

Une équipe actuelle de 8 personnes

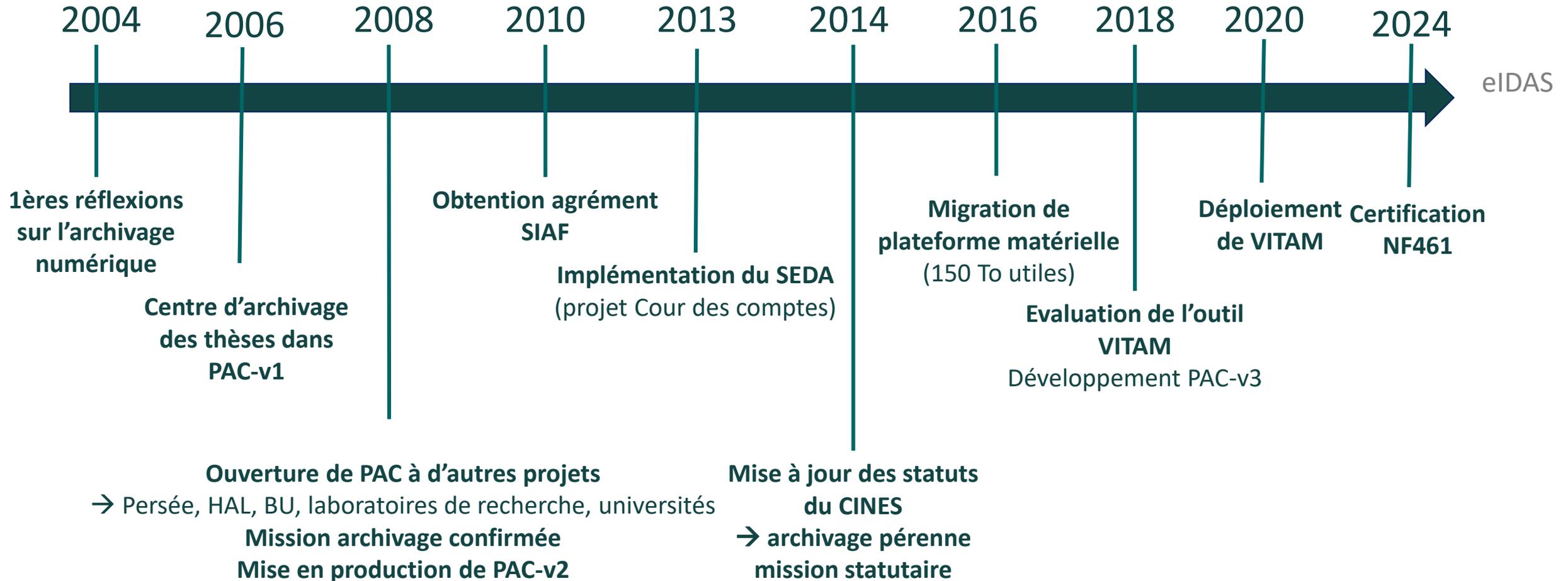
- 1 Responsable Département/projets;
- Mission archivage :
 - 1 eArchiviste;
 - 2 administrateurs applicatifs;
 - 1 développeurs.
- Activité Développement : 2 développeurs, 1 projet européen.

Activités

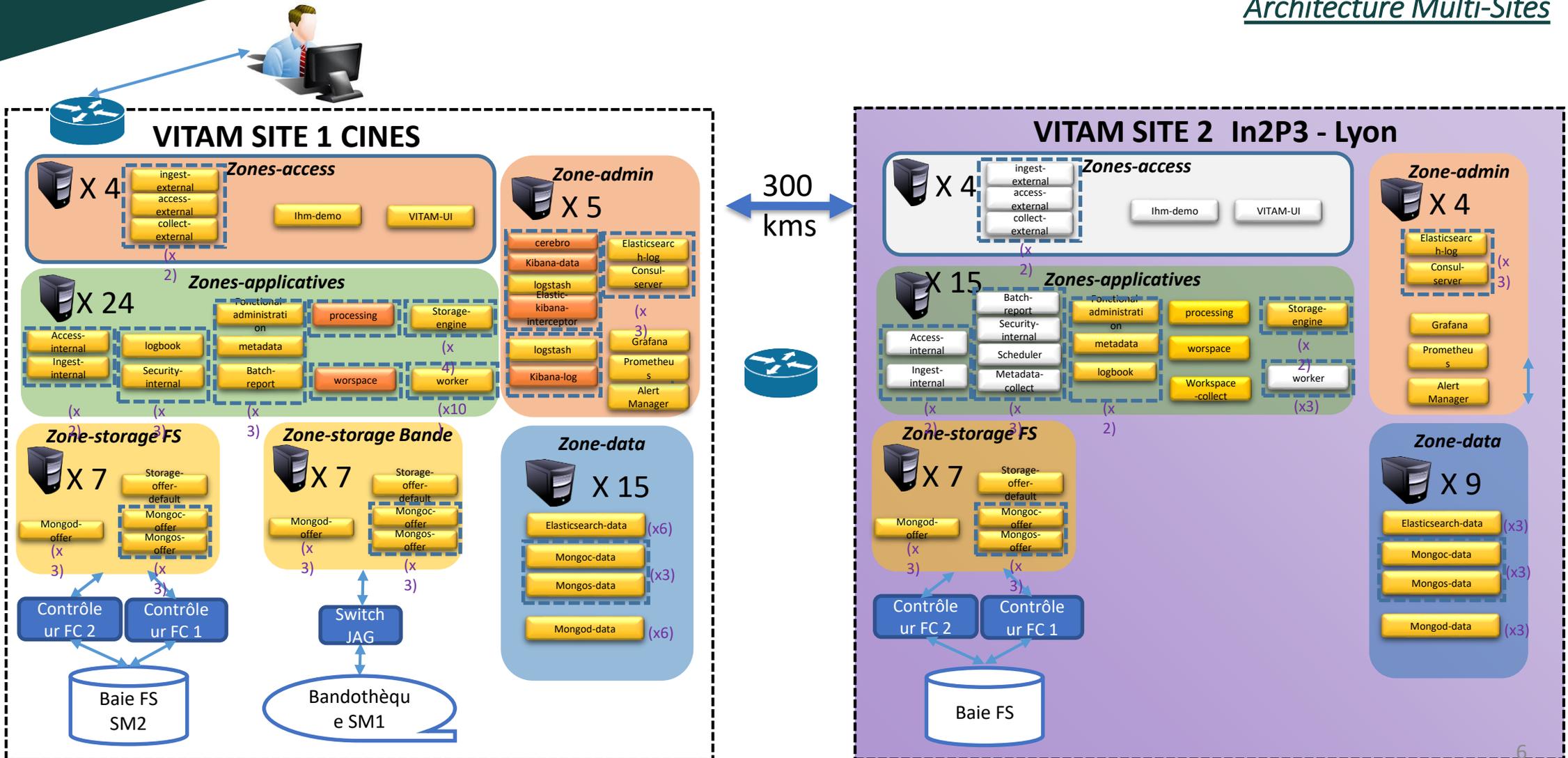
- Mission Archivage :
 - MCO de deux SAE : PACv3 et PACv4(base VITAM);
 - Développement archivage (Facile, validateur de format, BIR, PAC);
 - Accompagnement des services versant;
 - Processus Métier et Qualité/Agrément;
- Développement : ESGBU, Edari, RESER.

Modèle économique

- Dotations de l'état;
- Facturation de la prestation d'archivage;
- Partenariats.



- 2 SAE en production : PACv3 et PACv4 (VITAM);
- Mise en conformité de la nouvelle version de PAC basé sur VITAM v6.2;
 - 1 site principal (production et pre-production);
 - 1 site PRA secondaire au CC-IN2P3 synchronisé;
- Prévisionnel :
 - Certification NF461 : fin Novembre 2024 puis agrément de l'état;
 - Migration des archives de PACv3 vers PACv4 et mise en place d'un connecteur;
 - Prise en compte des projets supplémentaires en attente (12);

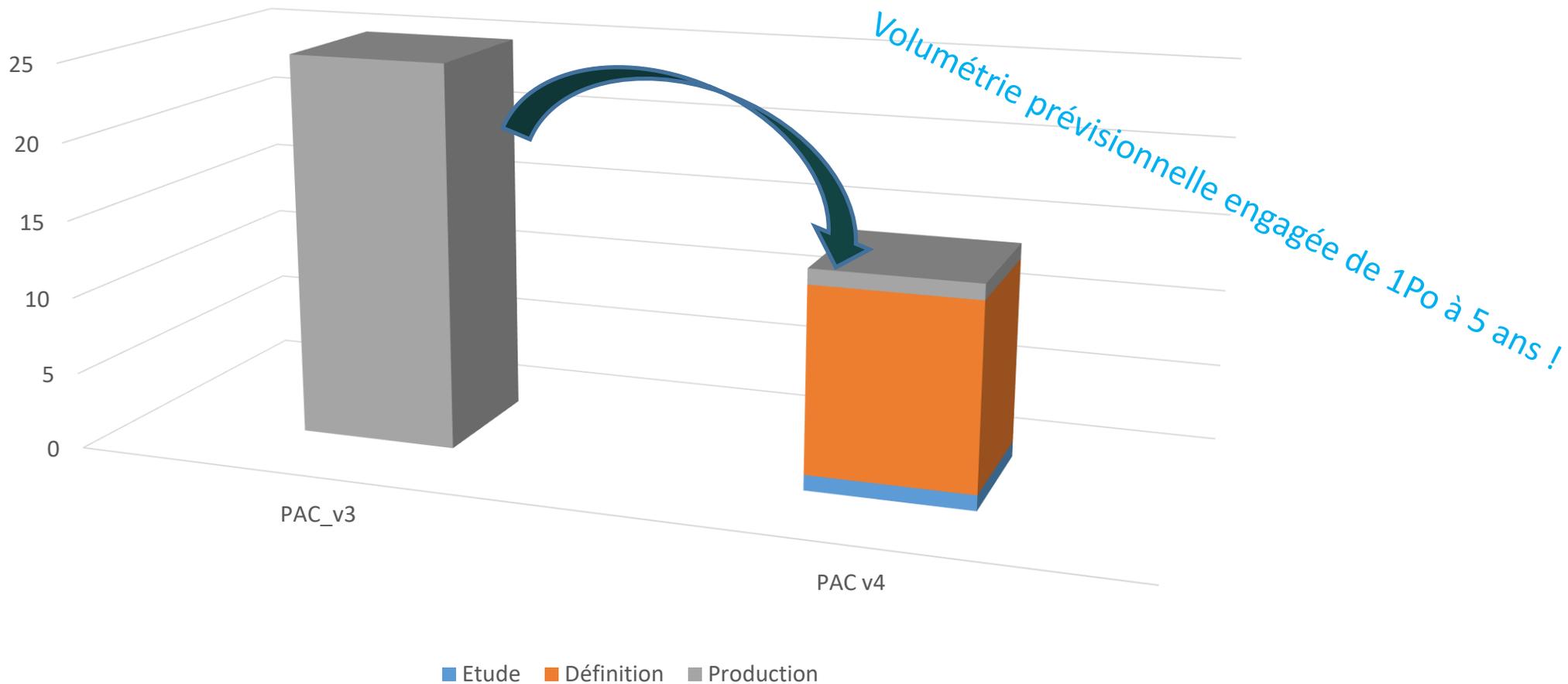


Capacité de Stockage et Extensibilité

	Site primaire (CINES)	Site Secondaire (CC-IN2P3)
serveurs	Dell PowerEdge R640	Hpe Proliant DL360
stockage	Dell EMC SCv3020	HPe MSA 2060 FC
Capacité utilisée	10 Tio HDD, 5 Tio SSD.	6.5 Tio HDD
Capacité utile	252 Tio HDD, 12 Tio SSD.	510 Tio HDD.
Extensibilité	800Tio utile HDD , 30Tio utile SSD	900 Tio utiles HDD
Total possible	1 Pio	1 Pio

Site 1 : stockage sobre et de grande masse sur 2 Robotiques IBM TS4500 partagées avec le HPC avec chacune plus de 40 Po Possible pour une faible consommation de 1500w.

40 projets d'archives en cours ou à venir



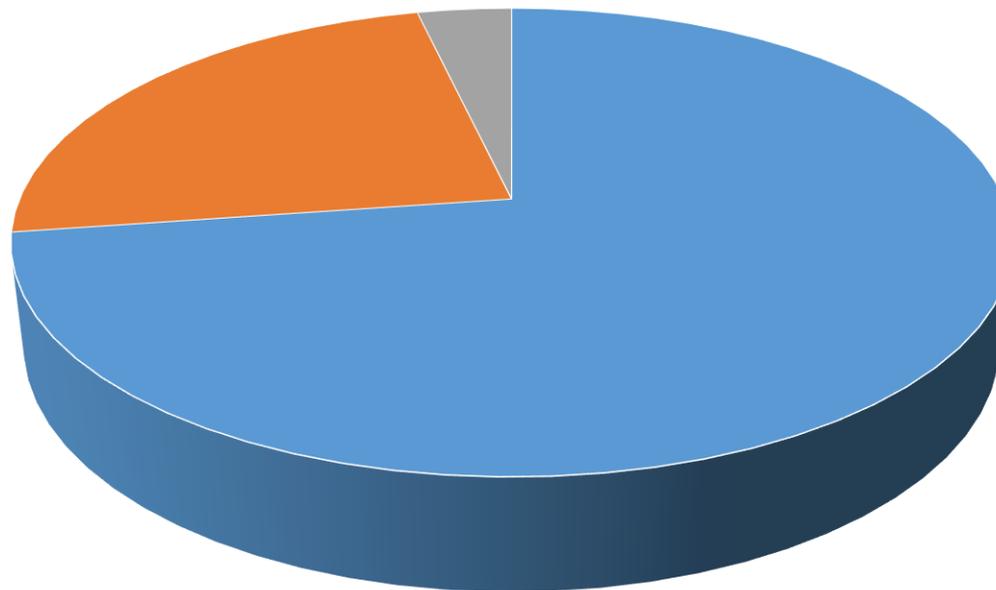
- Trois modes de financements pour l'archivage :
 - Facturation de la prestation : *Actuellement, principalement sur une base volumétrique;*
 - Dotation de l'état;
 - Partenariats : exemple du projet Ecol+.
- Travaux sur les coûts complets : *Périmètre important !*
 - Immobilisations : *Terrain, bâti, énergie, fluides,..*
 - Infrastructure et matériel : *Acquisitions, amortissement...*
 - Fonctionnement : *Ressources Humaine, prestation, maintenance, Qualité/sécurité, audit.....*

- Cadre réglementaire et normatif :
 - NF Z42-013/NF641, RGPD, ANSSI/sécurisation, Eiads, SEDA, valeur probante;
- Technologies de plus en plus complexes:
 - Solution VITAM (ElasticSearch, API rest, base mongoDb, etc..);
- Une compétence spécifique, longue à acquérir et volatile :
 - Coût important du turn-over (Earchiviste, processus métier, administrateur SAE et système, architecte, spécialistes robotiques, réseaux....);
- Gestion administratives :
 - gestion des conventions et de la facturation de plus de 30 projets;
- La planification de la préservation à un cout important :
 - Expertise format, BIR, processus métier, développements spécifiques (facile, validateur de formats)...

5 personnes : (5 DASD+1 admin sys+1 gestion globale)+ prestation

- la plateforme Montpellier incluant 5 ans de maintenance;
- la plateforme Lyon: incluant 5 ans de maintenance;
- les services mutualisés indispensables (DNS, mail, authentification, supervision...) une surcote de 30% sur la partie précédente
- 2 robotiques + maintenance + bandes pour le stockage. Usage de 10% pour l'archivage
- Hébergement site distant, fluide ...;

Répartition des coûts de l'archivage



■ RH ■ Infra ■ Audit/qualité/sécurité

Des coûts pour les déposants

- Financier : Tarification volumétrique;
- Mise en œuvre du projet d'archivages : plusieurs étapes...
- Connaissance du SEDA;
- Connaissance technique : utilisation des interfaces du SAE;
- Mise en œuvre/développement de l'interface technique;

Contexte	Problématique	Solution possible
Déposants	Manque de moyens technique, et archivistique pour les petites structures. Problématique de la coordination des jeux de données pour une même discipline	Service commun ayant la compétence métier, archivistique, et technique pour s'interfacer avec un SAE et élaborer des standard de description métier
Opérateur du SAE	Les spécificités techniques et les compétences métiers ont un coût élevé à acquérir et maintenir (turn-over, maintenance)	Augmenter la coopération (exemple projet VITAM, PIN...) sur le plus grand spectre des processus du SAE. (ex : archivage pérenne).
financement	Dépendance du modèle économique de l'établissement	Financement au plus haut niveau possible..

Constat

- Coût important du cadre normatif;
- Coût important de la planification de la préservation à long terme;
- Coût important du maintien des compétences spécifiques.

Leviers

- Coopération en terme d'expertise et de services..
- ...

Merci pour votre écoute



Voilà c'est fini ...

Les questions, remarques ou avis sont
les bienvenus.



CONTACT

Philippe PRAT

Responsable du DASD – CINES

prat@cines.fr