



La préservation numérique au CINES

Alexia de CASANOVE

Présentation Capitoul– 19 octobre 2017



Des exigences de plus en plus fortes autour des données...

Réutilisation

Open data

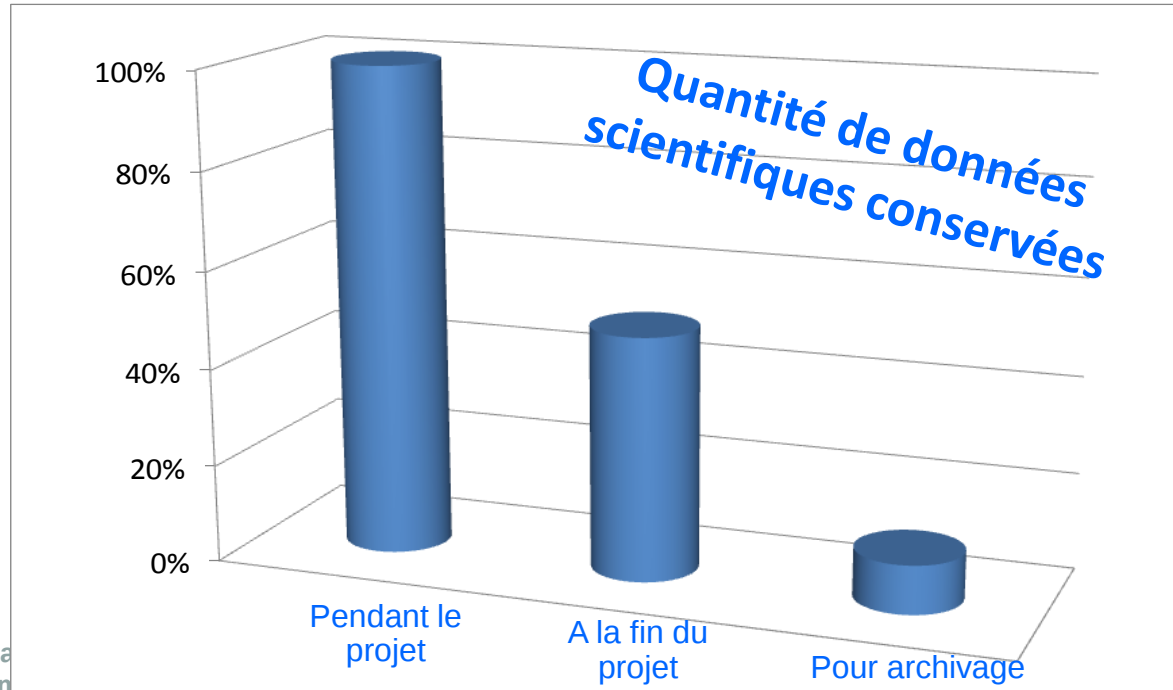
Big Data

Interopérabilité

Propriété des données



... qui nécessitent de s'interroger sur les moyens de pérennisation



Vos FILMS et VIDÉOS de famille **sauvegardés à vie** sur DVD !

**DVD DE FAMILLE: le spécialiste de la sauvegarde
de vos archives audiovisuelles familiales.**

TRANSFERT SUR DVD DE TOUS LES FORMATS
Super 8 - 8 mm - 9,5 mm - 16 mm
VHS - Hi8 - mini DV - Umatic - Betacam - Diapositives

TEST GRATUIT sur 3 minutes de film.
Prêt sur place d'une visionneuse pour repérage
et sélection de vos films.

SERVICE DE MONTAGE A LA CARTE
des meilleurs moments de vos films et vidéos.
TRANSFERT SUR CD AUDIO
de vos bandes, cassettes audio et disques vinyl.

Enlèvement gratuit sur Paris de vos films par le coursier DVD de Famille.

DVD DE FAMILLE

Depuis plus de 7 années d'existence, des milliers de clients ont confié à DVD DE FAMILLE leur patrimoine familial audiovisuel. DVD DE FAMILLE est régulièrement cité par les médias comme le prestataire vous garantissant l'expérience, la qualité, la rapidité, au juste prix.

57, RUE RENNEQUIN
75017 PARIS
Métro : Péreire

Ouvert du lundi au samedi
de 10 h à 19 h, sans interruption.

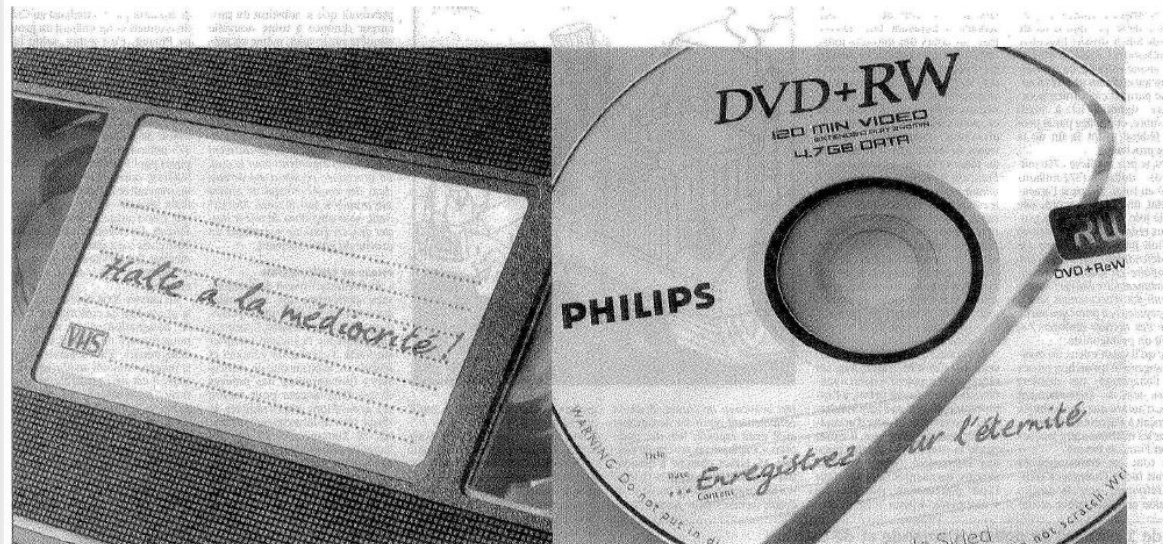
2, RUE DE LONGCHAMP
92200 NEUILLY / SEINE
Métro : Pont de Neuilly

Ouvert du lundi au vendredi
de 10 h à 13 h, et de 14 h à 19 h.

Devis immédiat par téléphone : 01 44 05 02 22

www.dvddefamille.com

NE PAS JETER SUR LA VOIE PUBLIQUE



Pour ne plus voir la qualité de vos enregistrements sur VHS disparaître avec le temps, entrez, pour moins de 500 euros, dans le monde du DVD Recorder. Avec la nouvelle gamme de lecteurs-enregistreurs DVD Philips simples d'utilisation et accessibles à tous, vous êtes assurés de **conserver pour l'éternité des enregistrements parfaits et inaltérables.**

www.philips.com/dvdr

Le Monde, 13 décembre 2003



Une disparition (généralisée) plus que probable

Des données coûteuses à produire

Difficiles à pérenniser :

- Difficultés de compréhension « intellectuelle » du contenu
- Ignorance de l'existence
- Soucis techniques (format / codage / matériel ...)
- ...



Peu d'intérêt des tutelles pour la préservation :

- Travail non valorisé dans la carrière des chercheurs
- Coûts généralement non intégrés dans les projets
- Une avancée avec le « data management plan » du programme H2020



SAUVEGARDE



Sauvegarde VS Sauvegarde sécurisée VS Archivage



Sauvegarde VS Sauvegarde sécurisée VS Archivage

ARCHIVAGE



- ✓ Respect de l'intégrité et de l'authenticité
 - ✓ Garantie de lisibilité
 - ✓ Garantie de compréhension
- Sur le moyen et long terme

Les enjeux de l'archivage numérique

• Triple engagement :



Préservation de l'intégrité

SUPPORT : Checksum + RAID +
contrôles CRC + VEILLE + MIGRATION
PHYSIQUE + copie sur SITE DISTANT

Préservation de la
capacité à
comprendre le
contenu du fichier

Précautions à prendre		Précautions à prendre	
	Température de lavage La température de lavage recommandée est indiquée en chiffres au centre du logo : 90°, 60°, 40°, 35° etc.		Ne pas laver à l'eau Un nettoyage à sec est recommandé.
	Laver à la main L'article ne doit pas être lavé en machine ni essoré. Température maximum de 40°.		Ne pas sécher en tambour L'article ne doit pas être mis dans un sèche-linge.
	Peut être passé au chlore (eau de javel)		Ne pas utiliser de chlore (eau de javel)
	Température de repassage Le logo peut comporter de 1 à 3 points en son centre. - 1 point : chaleur faible (fibres artificielles) - 2 points : chaleur moyenne (laine, soie) - 3 points : chaleur forte (coton, lin)		Ne pas repasser Concerne notamment certaines fibres synthétiques qui craignent la chaleur comme par exemple le Nylon.

Préservation de la lisibilité

Ancien format

Nouveau format



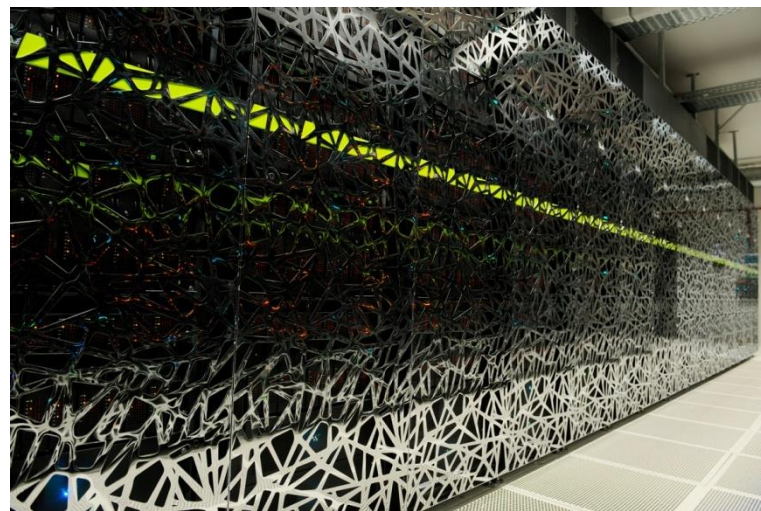
ENVIRONNEMENT MATERIEL : VEILLE
TECHNO et ANTICIPATION
ENVIRONNEMENT LOGICIEL : privilégier
les FORMATS DURABLES + MIGRATION
LOGIQUE (→ validation préalable de format)

METADONNÉES DESCRIPTIVES +
IDENTIFICATION UNIQUE et
PERENNE

L'opérateur ESR de mutualisation pour l'archivage

Centre Informatique National de l'Enseignement Supérieur

- Basé à Montpellier (Hérault, France)
- EPA créé en 1999, succédant au CNUSC – créé en 1980
- Placé sous la tutelle du Ministère chargé de l'Enseignement Supérieur et de la Recherche
- Missions
 - Calcul numérique intensif
 - Archivage pérenne de données électroniques
 - Hébergement de matériels informatiques à vocation nationale
- Plus d'informations : <http://www.cines.fr/>



La machine " Occigen" du CINES

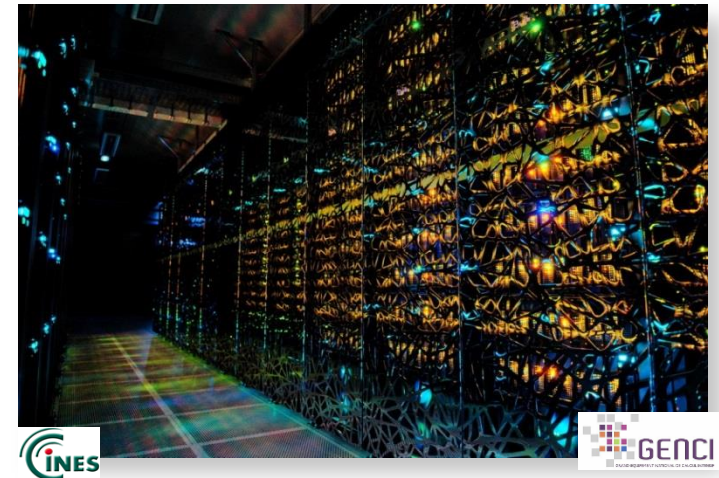
Infrastructures sécurisées

- 5 salles machines : 1 400 m²
- Locaux techniques : 2 000 m²
- Accès physiques et logiques sécurisés
- 2 lignes ERDF pour un total de 12,5 MW + fabrication autonome d'eau & air froids
- Données en double alimentation ondulée et sécurisée par groupe électrogène
- Copies et sauvegardes dans des salles distinctes + copie à distance



Ressources

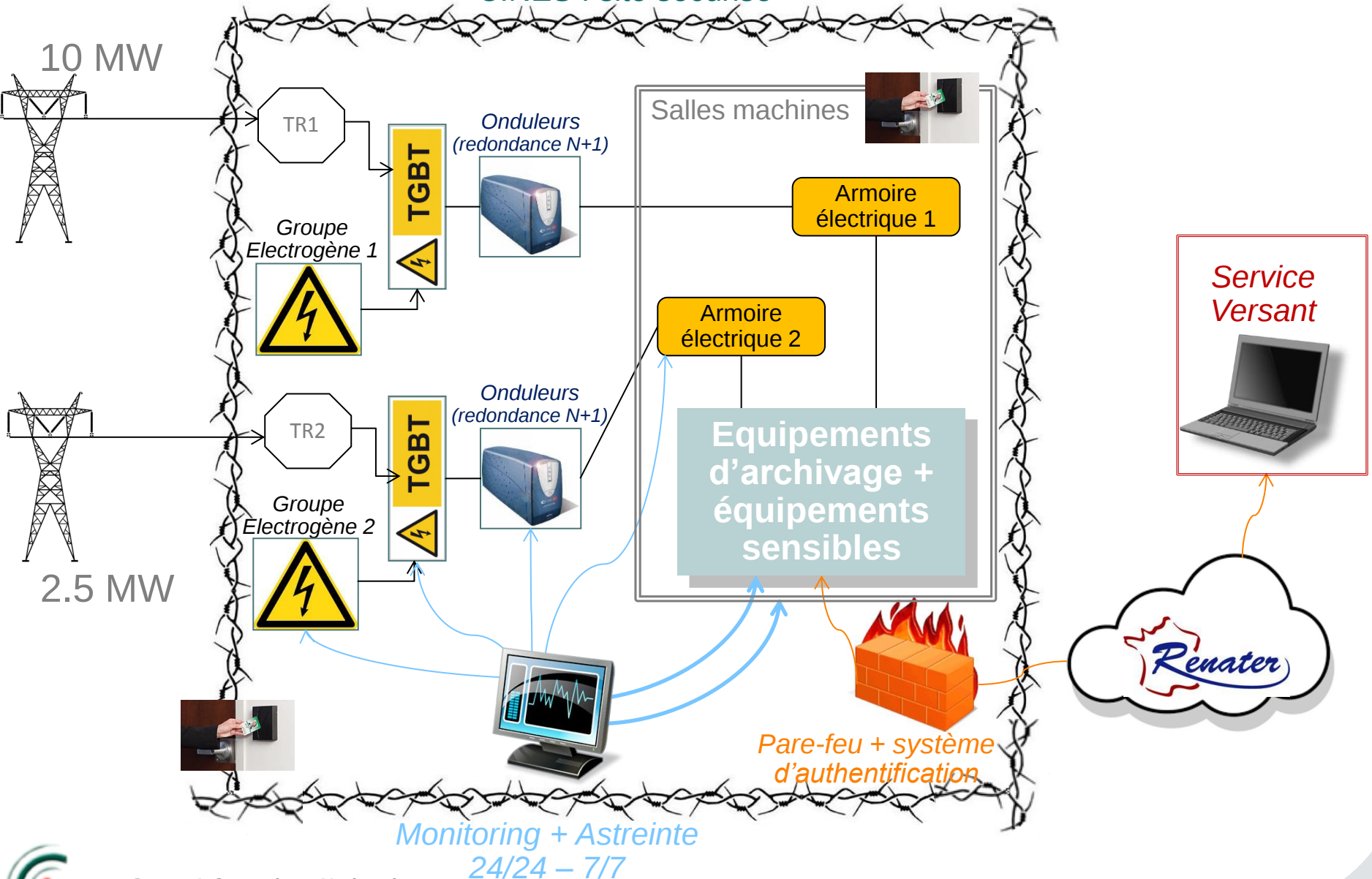
- Un supercalculateur de niveau mondial
- Capacités de stockage de plusieurs PetaOctets
- Des accès réseau performants
- Des équipes d'experts
- Une surveillance 24h/24 – 7j/7 avec astreinte



Le supercalculateur " Occigen "

Capitoul– 19 octobre 2017

CINES : site sécurisé



Des solutions techniques et organisationnelles

Nos solutions sont :



mutualisées



sécurisées



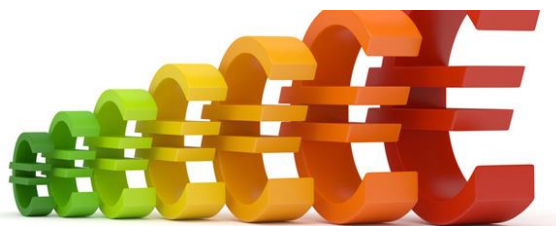
personnalisées



performantes



standardisées



économiques



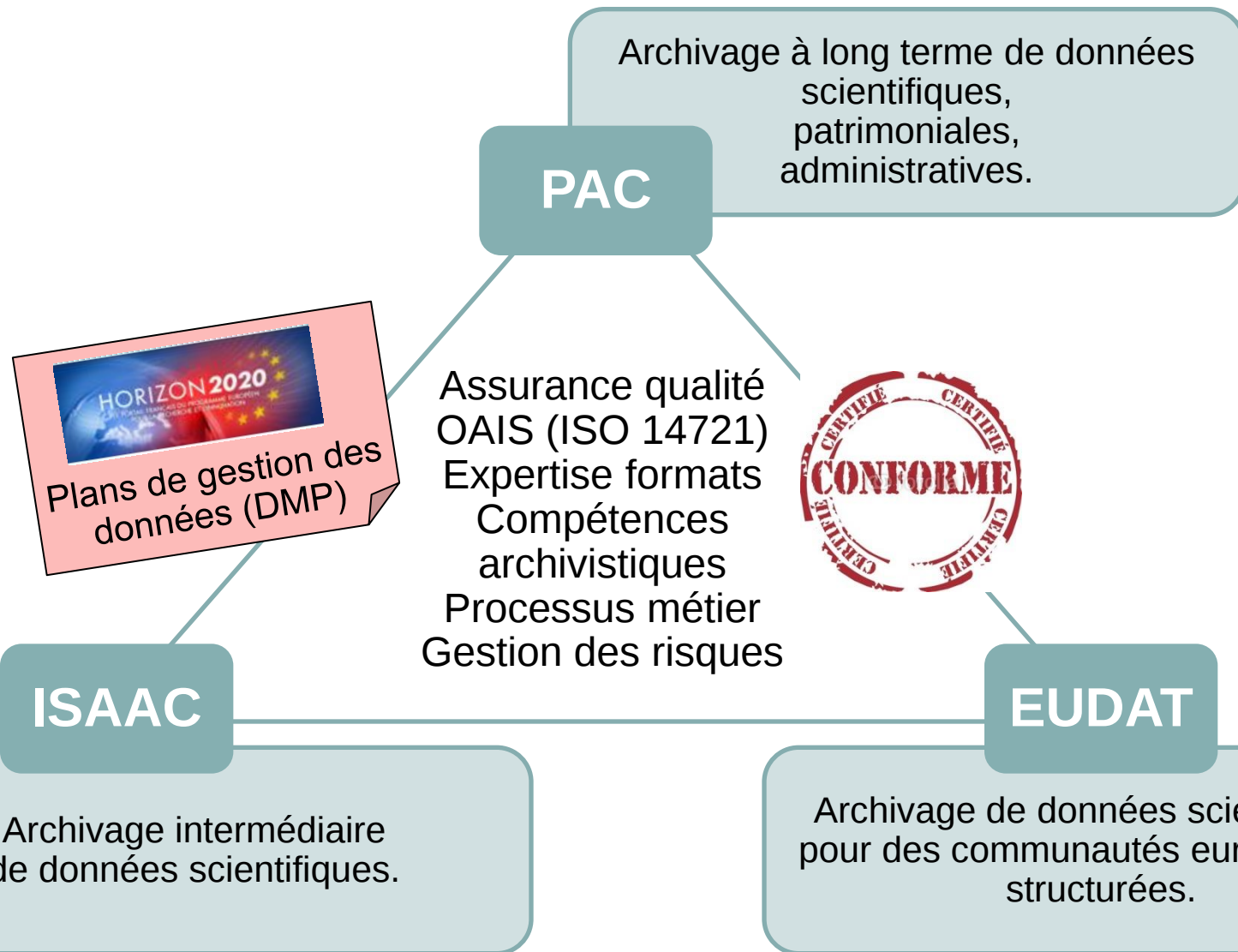
certifiées

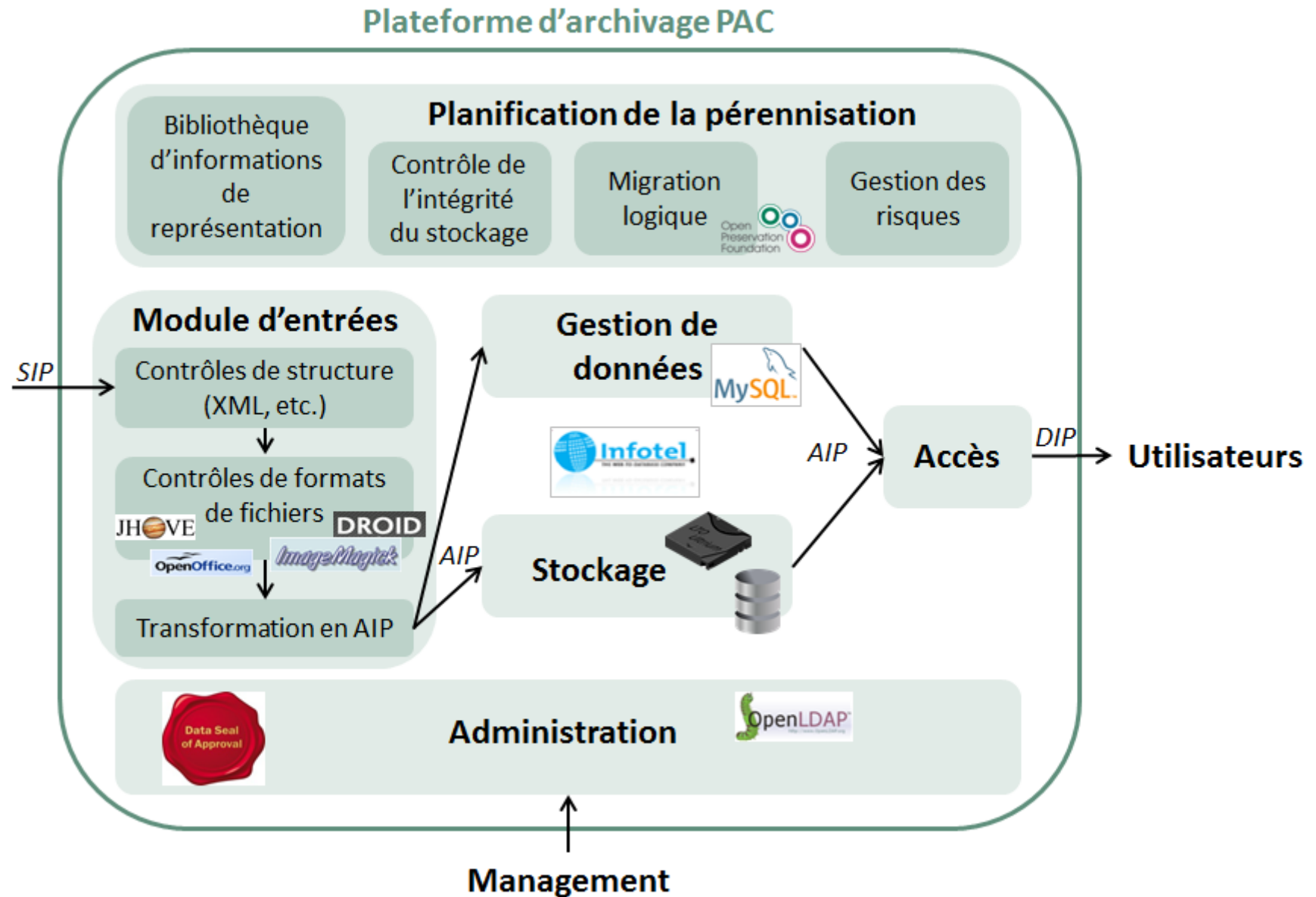


complémentaires
des solutions de
diffusion

→ CINES = unique opérateur mandaté par le ministère pour la pérennisation des données produites par la communauté ESR

Les infrastructures d'archivage du CINES





Les différents niveaux de métadonnées dans PAC

Fonds
ou
Projet
d'archives

PPDI.XML

Document

Sip.xml ou
ArchiveTransfer.xml /
AIP.XML

Metadonnees_
metier.XML

Bibliothèque
d'Informations de
Représentation

Spécifications des
formats des
fichiers archivés

Schémas XSD ou
DTD des fichiers
XML archivés

Profils d'archivage

PPDI

Conventions
d'archivage

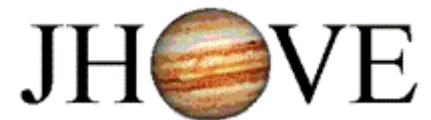
Pour permettre le contrôle de la qualité d'un fichier, celui-ci doit être dans un format identifié et vérifiable :

- Format publié ; ex. WAVE, SVG ;
- Format largement utilisé ; ex. XML, MPEG4 ;
- Format normalisé si possible ; ex. PDF (ISO 32000-1:2008), PNG (ISO 15948:2004).

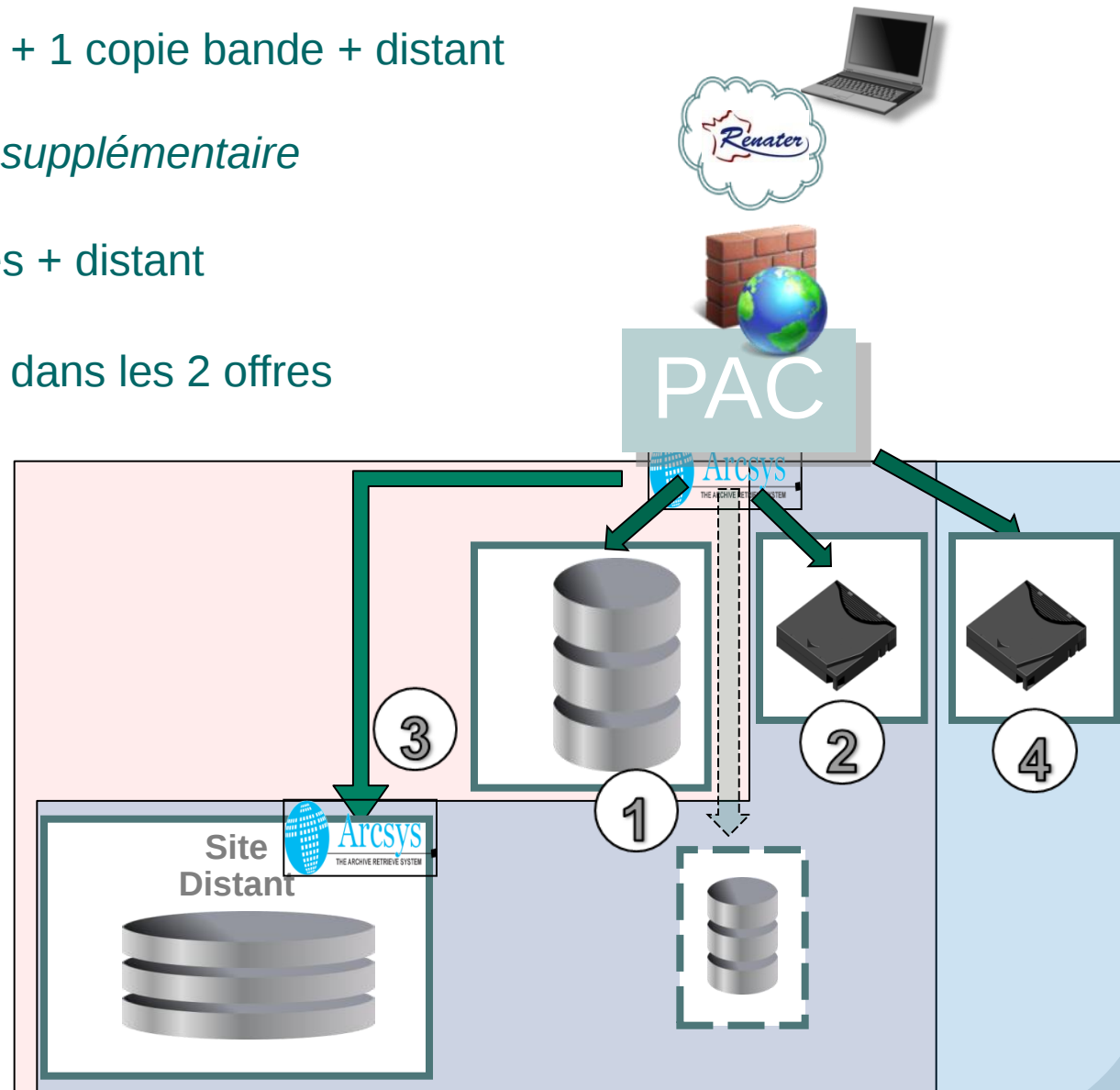
Pour pouvoir être lisibles dans le temps, et convertibles, les fichiers doivent respecter à la lettre les **spécifications** de leur format.

Les outils libres Jhove, ImageMagick, DROID, ODF Validator permettent une identification, validation et caractérisation des formats.

Type	Format
Texte	PDF, TXT, XML, ODT, TEI
Image	GIF, JPEG, TIFF, PNG, SVG, JPEG2000, GéoTIFF
Audio	WAV, AIFF, AAC, VORBIS
Vidéo	MPEG4, THEORA, MKV

The logo for JHOVE, featuring the word "JHOVE" in a serif font. The letter "O" is replaced by a detailed image of the planet Jupiter.The logo for ImageMagick, featuring the word "ImageMagick" in a stylized, italicized, blue font with a white outline.The logo for DROID, featuring the word "DROID" in a bold, white, sans-serif font on a dark, textured background.

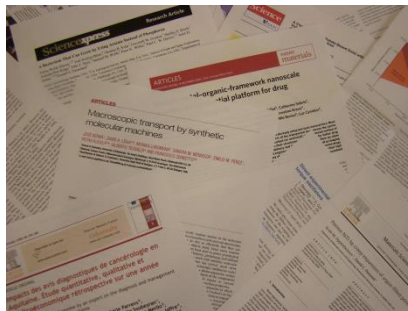
- Niveau 1 : 1 copie disque + 1 copie bande + distant
+ *en option* : 1 copie disque supplémentaire
- Niveau 2 : 2 copies bandes + distant
- Une copie distante native dans les 2 offres



Les candidats à l'archivage



**Thèses de doctorat
françaises**



Publications scientifiques



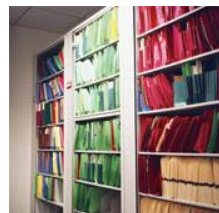
**Documents
patrimoniaux
numérisés**



**Vidéothèques et
banques d'images**



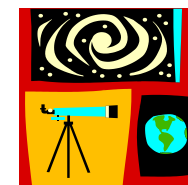
Rapports



**Données de
gestion**



Résultats de calculs



Observations

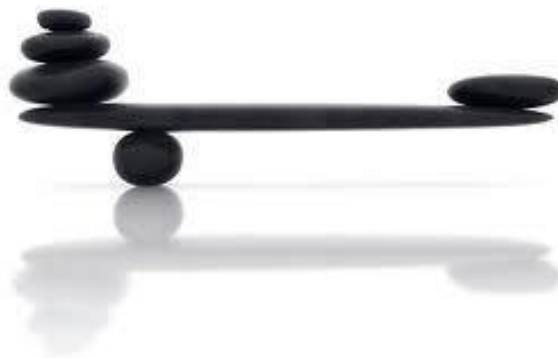
Un projet d'archivage au CINES, c'est... ?

- L'archivage de données produites par les organismes de l'ESR :
 - ✓ **validées** (archivage courant, intermédiaire et définitif)
 - ✓ **sélectionnées**
 - ✓ **documentées** (jeu minimal de métadonnées : Dublin Core / SEDA)



Quel travail de curation sur les données pour l'archivage ?

- *adapté à la durée de conservation des données*
- *et au niveau de services demandé : intégrité, lisibilité, intelligibilité*



Un projet d'archivage au CINES, c'est... ?

- **Une équipe-projet dédiée :**

- ✓ Un référent-projet informatique et un archiviste côté CINES
- ✓ Un référent-projet côté Service Versant
- ✓ Des développements informatiques à prévoir :
interfaçage avec la plateforme



- **Un partenariat encadré :**

- ✓ Lettre d'intention
- ✓ Convention d'archivage
- ✓ Tarification au Téraoctet utile archivé et en fonction du niveau de service



Fonctionnement de la plateforme PAC

Plateforme d'archivage (agrée SIAF + DSA + ISO 16363)

1. Réception

- Authentification LDAP
- Transfert du SIP
- Envoi AR



2. Contrôle Qualité

- Contrôle checksum-prod
- Contrôle forme du SIP
- Contrôle format des fichiers



3. Création de l'AIP

- Génération checksum-SAE
 - Date d'archivage
 - PID (ARK, DOI, Handle...)
 - [Conversion fichiers dans Format d'archivage]
- ...

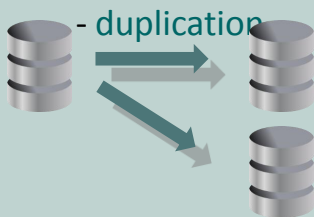
4. Traitements complémentaires

- Récupération d'informations associées (Info de représentation...)

5. Stockage de l'AIP en plusieurs exemplaires + référencement

7. Migration physique :

- rafraîchissement de support



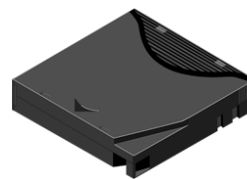
8. Migration logique :



Salle machine 1



Salle machine 2



Salle machine 3



Site distant (> 300km)



6. Vérification périodique de tous les exemplaires archivés :

- Vérification checksum-SAE
- Validation bijectivité AIP/référencement
- Validation non corruption du référentiel



Exemple : « l'archivage des données satellitaires Géosud »

Type de données concernées :



- Images satellitaires brutes et orthorectifiées
- Couverture annuelle nationale et régionale, acquisitions en pied d'antenne (SPOT6/7)
- Éléments de volumétrie : environ 10 To/an

Niveau de service assuré :

- Archivage pérenne avec conservation sur le long terme
- Différents niveaux de métadonnées de description : Dublin Core, ISO 19115
- Contrôle de la validité des formats des fichiers : GeoTIFF, JPEG2000, XML



Intérêt de l'archivage : pouvoir mesurer les évolutions des territoires

Problème de conformité à la norme GeoTIFF (système de projection Lambert 93 utilisé pour les images orthorectifiées absent de la norme).

Solution retenue : archivage des images brutes en plus



Cour des comptes



- Nos partenaires



Centre Informatique National
de l'Enseignement Supérieur

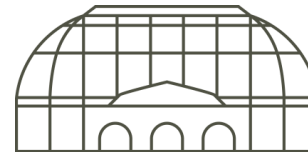
- Nos partenaires



UNIVERZITA KARLOVA



MUSÉUM
NATIONAL D'HISTOIRE NATURELLE



BG
BM | Botanischer Garten &
Botanisches Museum
Berlin



Royal
Botanic Garden
Edinburgh



Jardin botanique
Meise



Merci, des questions ?



Plus d'informations : <http://www.cines.fr/archivage/>

CONTACT

Marion MASSOL

Responsable Département Archivage et Diffusion – CINES

massol@cines.fr