



# La préservation numérique au CINES

Lorène BECHARD

*Colloque DataBFC – 14 novembre 2017*



# Les enjeux de l'archivage numérique

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

**Réutilisation**

**Open data**

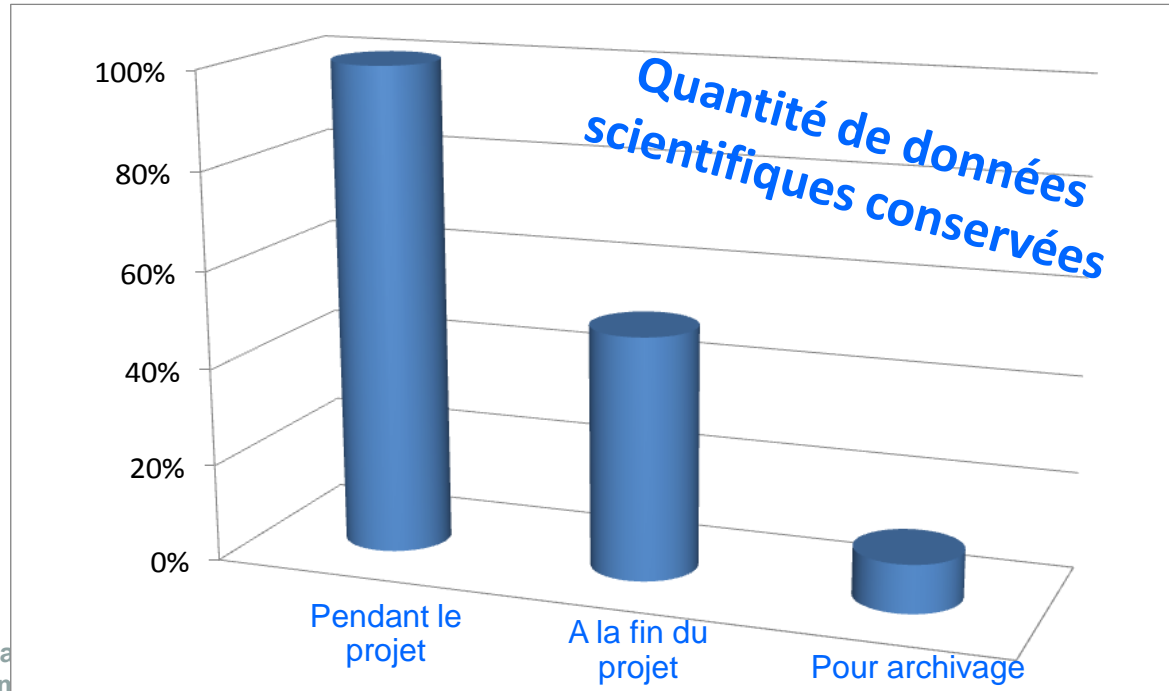
**Big Data**

**Interopérabilité**

**Propriété des données**



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



# Il y a ce qu'on nous promet...

## 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.

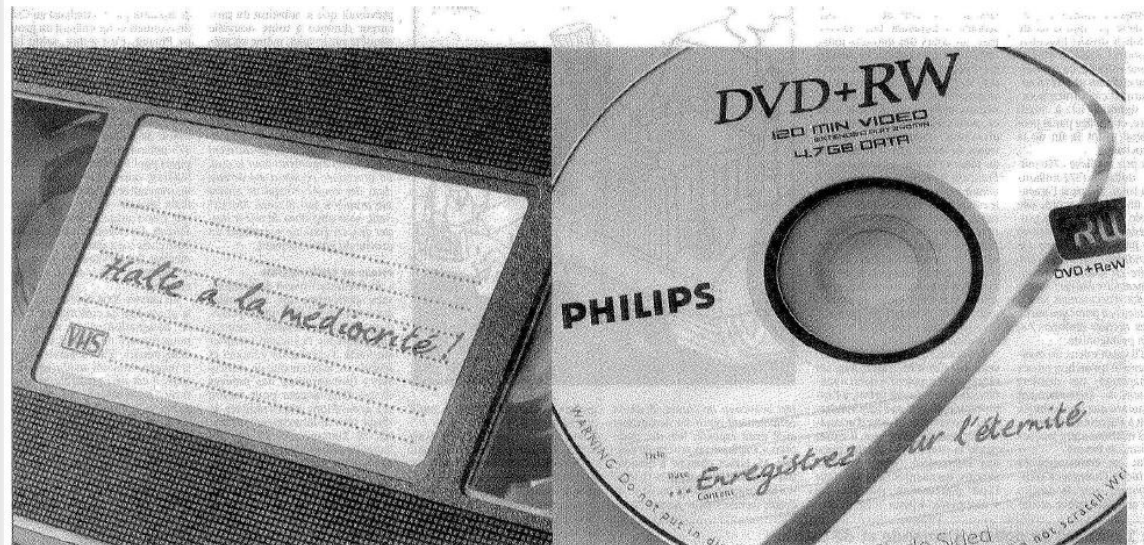
2, RUE DE LONGCHAMP  
92200 NEUILLY / SEINE  
Métro : Pont de Neuilly

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](http://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](http://www.philips.com/dvdr)

Le Monde, 13 décembre 2003



# Les défis, orientations et choix pour l'archivage au CINES



« La NASA avait oublié depuis longtemps ce logiciel ou les programmeurs qui le connaissaient étaient morts » [Joseph MILLER]



# 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

## • 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	
	<b>Température de lavage</b> La température de lavage recommandée est indiquée en chiffres au centre du logo : 90°, 60°, 40°, 35° etc.		<b>Ne pas laver à l'eau</b> Un nettoyage à sec est recommandé.
	<b>Laver à la main</b> L'article ne doit pas être lavé en machine ni essoré. Température maximum de 40°.		<b>Ne pas sécher en tambour</b> L'article ne doit pas être mis dans un sèche-linge.
	<b>Peut être passé au chlore</b> (eau de javel)		<b>Ne pas utiliser de chlore</b> (eau de javel)
	<b>Température de repassage</b> 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)		<b>Ne pas repasser</b> Concerne notamment certaines fibres synthétiques qui craignent la chaleur comme par exemple le Nylon.

## Préservation de la lisibilité

Ancien format

Nouveau format



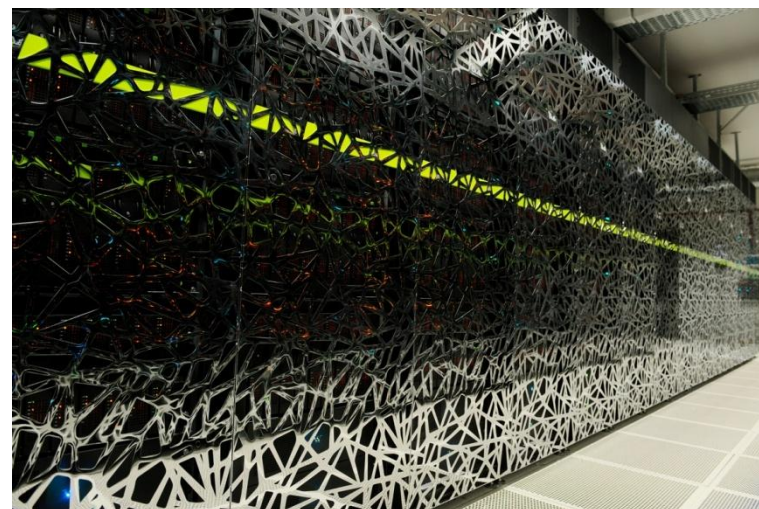
ENVIRONNEMENT MATERIEL : VEILLE  
TECHNO et ANTICIPATION  
ENVIRONNEMENT LOGICIEL : privilégier  
les FORMATS DURABLES + MIGRATION  
LOGIQUE (→ validation préalable de format)

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

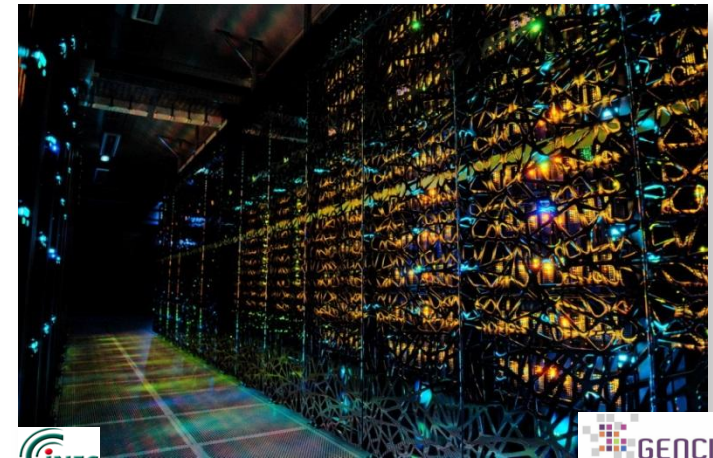
## Infrastructures sécurisées

- 5 salles machines : 1 400 m<sup>2</sup>
- Locaux techniques : 2 000 m<sup>2</sup>
- 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



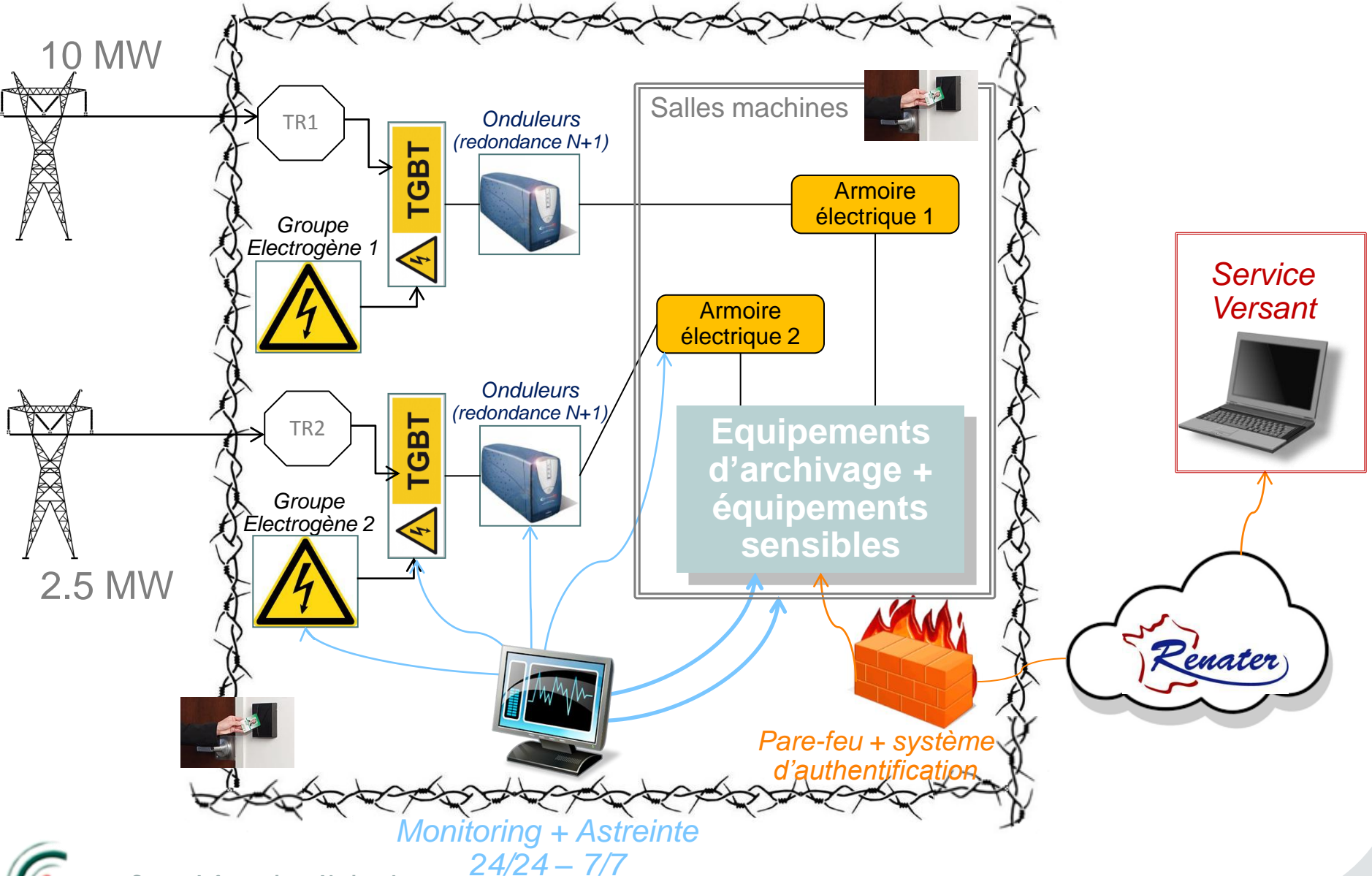
## Ressources

- 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 "*

## CINES : site sécurisé



# Des solutions techniques et organisationnelles

## Nos solutions sont :



mutualisées



sécurisées



personnalisées



performantes



standardisées



économiques



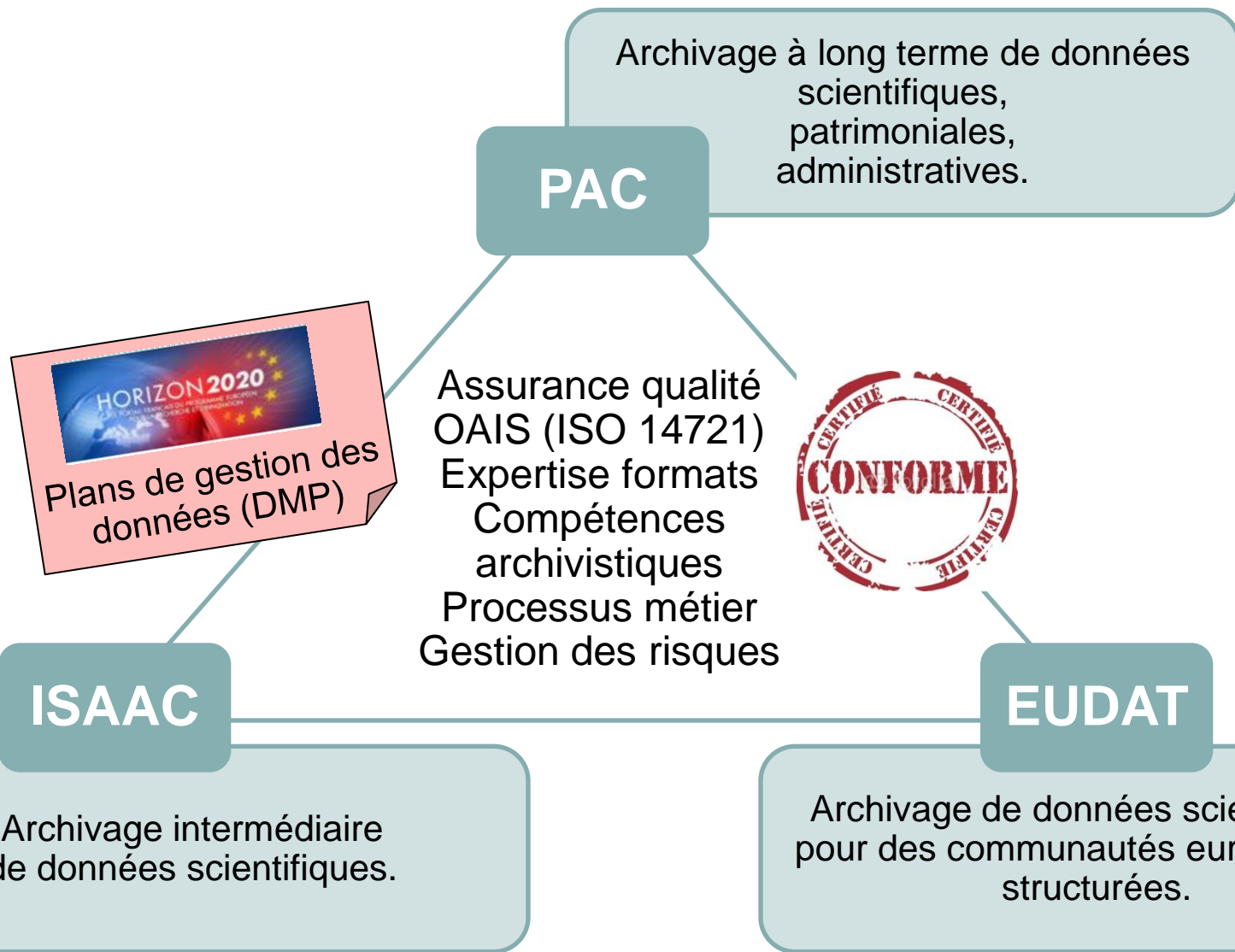
certifiées



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





**Thèses de doctorat  
françaises**



**Publications scientifiques**



**Documents  
patrimoniaux  
numérisés**



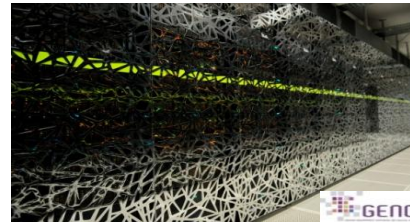
**Vidéothèques et  
banques d'images**



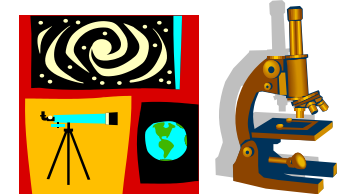
**Rapports**



**Données de  
gestion**



**Résultats de calculs**



**Observations**

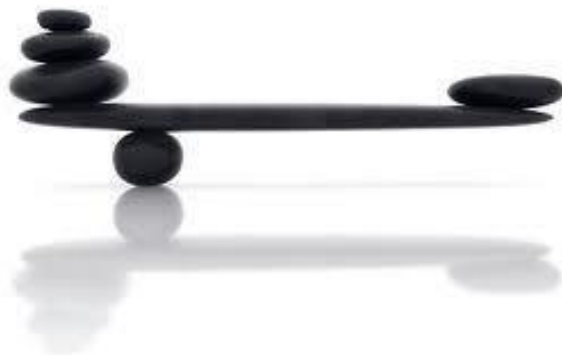


- **L'archivage de données produites par les organismes de l'ESR :**
  - ✓ **validées** (archivage courant, intermédiaire et définitif)
  - ✓ **sélectionnées**
  - ✓ **documentées** (jeu minimal de métadonnées : Dublin Core / SEDA)



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...)

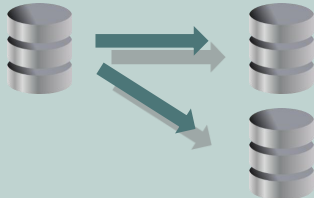
### 5. Stockage de l'AIP en plusieurs exemplaires + référencement

### 7. Migration physique :

- rafraîchissement de support



- duplication



- ré-empaquetage



### 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



# Ils archivent au CINES...



Cour des comptes



- Nos partenaires



Huma-Num



SciencesPo.



revues.org



Maison méditerranéenne des sciences de l'homme



CINES Centre Informatique National de l'Enseignement Supérieur

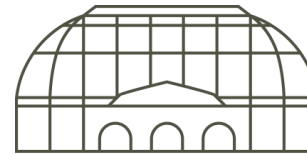
- Nos partenaires



UNIVERZITA KARLOVA



MUSÉUM NATIONAL D'HISTOIRE NATURELLE



BG | 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](mailto:massol@cines.fr)