



# L'archivage électronique au CINES

Marion MASSOL (CINES)  
[marion.massol@cines.fr](mailto:marion.massol@cines.fr)

Marseille Workshop on Scientific Data Preservation – 21 novembre 2012





1. Qu'est-ce que l'Archivage Electronique ?
2. Le CINES : missions et stratégie
3. PAC : une solution nationale d'archivage
4. EUDAT : une grille européenne de données



# Qu'est-ce que l'archivage électronique ?



L'archivage pérenne des documents électroniques consiste à conserver le document et l'information qu'il contient :

- ✓ Dans son *aspect physique* comme dans son *aspect intellectuel*,
- ✓ Sur le très long terme soit **30 ans et au-delà**,
- ✓ De manière à pouvoir **le rendre accessible et compréhensible**.

# Qu'est-ce que l'archivage électronique ?

## Les défis, orientations et choix pour l'archivage électronique

Voici un document  
que j'ai créé en  
1998...



De quoi s'agit-il déjà ?  
Est-ce bien ce qui est  
indiqué sur la  
disquette ?

METADONNEES DESCRIPTIVES + IDENTIFICATION  
UNIQUE et PERENNE



La disquette est-elle  
toujours en bon état ?

SUPPORT : VEILLE +  
MIGRATION PHYSIQUE



Mon portable, acheté  
en 2009, n'a pas de  
lecteur de disquette...

ENVIRONNEMENT MATERIEL :  
VEILLE TECHNO et ANTICIPATION



Ça marche ! Mais j'ai  
perdu toute ma mise  
en forme...

INTEGRITE AUTHENTICITE



J'ai trouvé le logiciel,  
mais puis-je l'installer  
et l'utiliser sous  
Windows 7?

SYSTEME D'EXPLOITATION

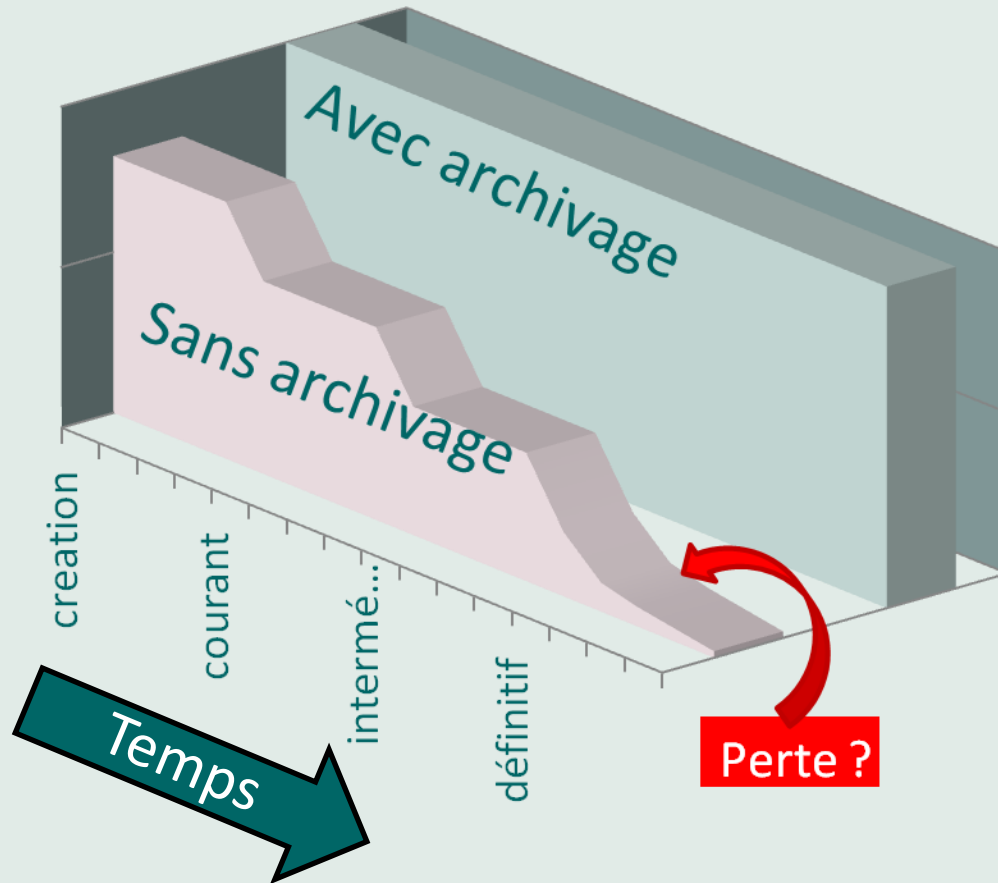


J'ai créé ce document avec Claris  
Works. Comment retrouver ce  
logiciel ? Quel est le format du  
document ?

ENVIRONNEMENT LOGICIEL : privilégier  
les FORMATS DURABLES + MIGRATION  
LOGIQUE

# Qu'est-ce que l'archivage électronique ?

## La dégradation de l'information au cours du temps



Les risques portent sur :

- Compréhension
- Intégrité
- Exploitation
- Valorisation

Mise en place de procédures d'assurance qualité

- Métadonnées
- Formats
- Stockage
- Info de représentation
- Communauté structurée
- Veille technologique.

## Depuis 2004, le CINES a une mission nationale d'archivage du patrimoine scientifique.

- Arrêté du 7 août 2006 relatif aux modalités de dépôt, de signalement, de reproduction, de diffusion et de conservation des thèses ou des travaux présentés en soutenance en vue d'un doctorat ;
- Lettre de cadrage du 12 février 2008 recentrant les activités du CINES autour de deux missions stratégiques : le calcul intensif et l'archivage pérenne.

## Objectifs : la mise en place de solutions mutualisées

- Performantes pour la conservation à moyen et long terme du patrimoine numérique des établissements
- Economiques et sécurisées

## Les données concernées sont :

- Les données scientifiques – résultats d'observations ou de calcul ;
- Les données pédagogiques, publications, etc. ;
- Les données administratives, archives intermédiaires.



Projets d'un an renouvelable

Données stockées dans un labo, hébergeur, centre de calcul...

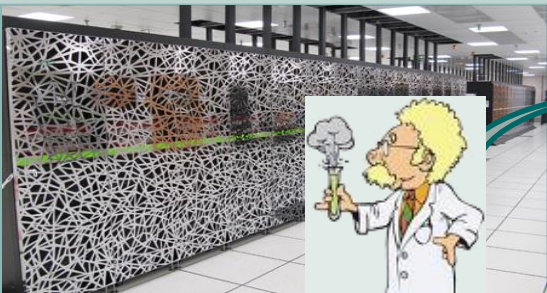


Transfert des données pertinentes  
→ Valeur ajoutée

Données produites dans un contexte **national** (droits spécifiques...)

**ISAAC** :  
Espace de conservation pour la durée du projet (3 à 5 ans max.)

**PAC** :  
Espace d'archivage pérenne



Données échangées au sein d'une **communauté européenne**

**EUDAT** :  
Grille européenne de données scientifiques préservées

SARA

JUELICH

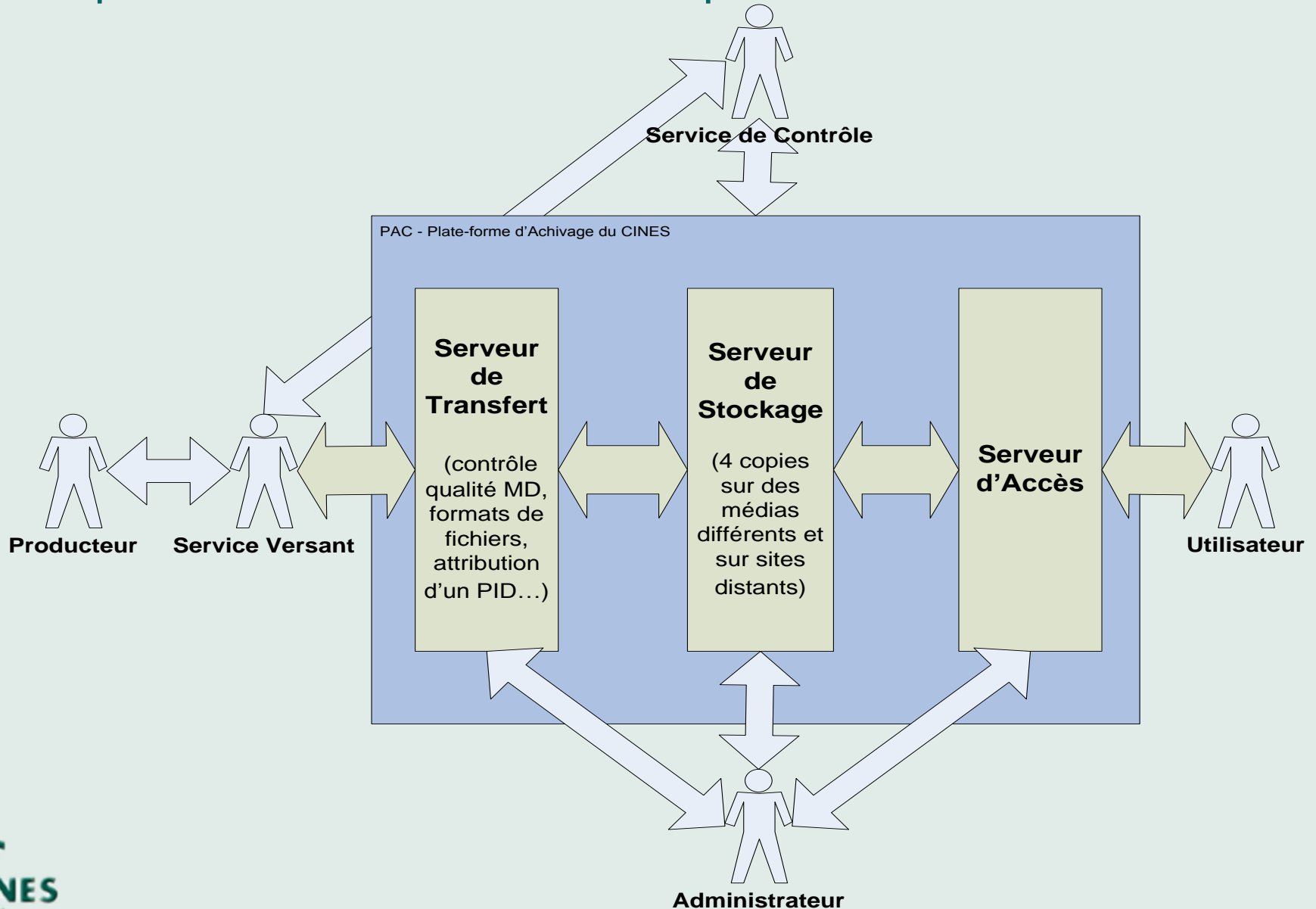
...

RZG

CINES: 500 To au début

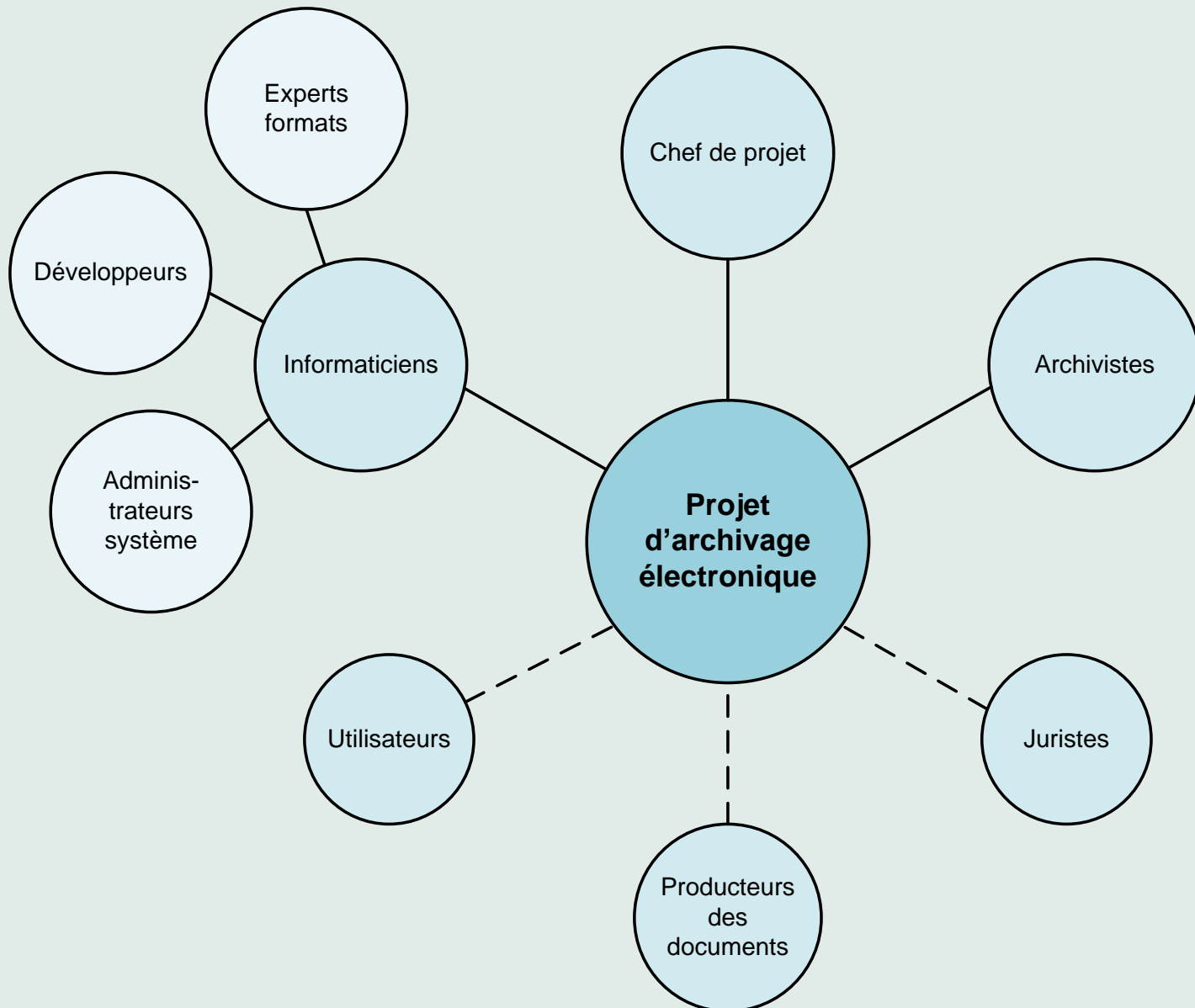


- Une plateforme mutualisée en exploitation :





- Une équipe dédiée de 12 ETP : 1 chef de projet, 9 ingénieurs, 1 archiviste, 2 techniciens



# PAC : une solution nationale d'archivage

## PAC : Ses utilisateurs



COUR DES COMPTES



- Projet de 3 ans : octobre 2011 – septembre 2014
- 25 partenaires de 15 pays européens :



# EUDAT : une grille européenne de données

## Un projet avec de nombreux partenaires

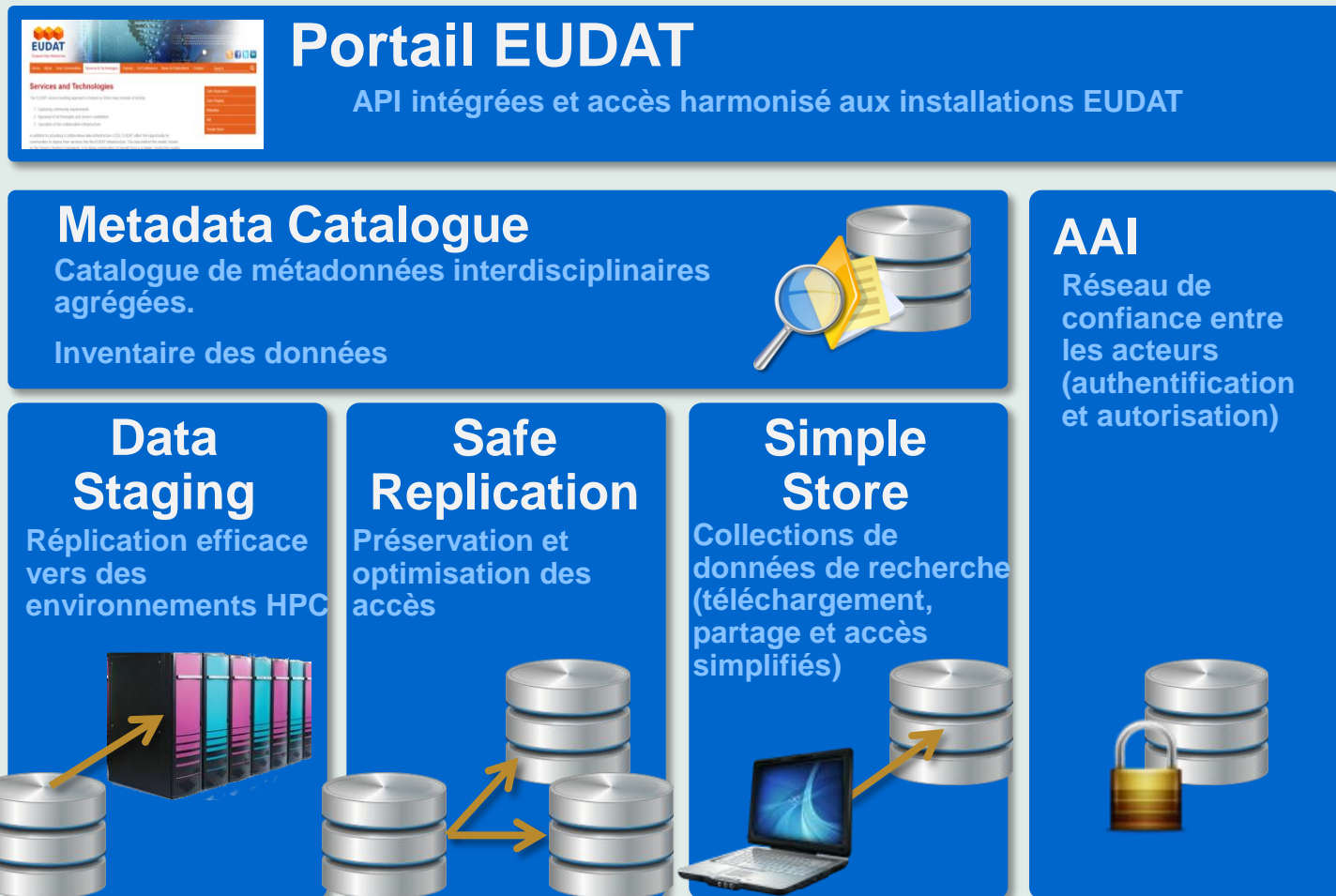
- Des échanges entre communautés et centres de calcul:



# EUDAT : une grille européenne de données

## Une infrastructure modulaire en cours de création

- Les principaux modules de l'infrastructure collaborative de données :

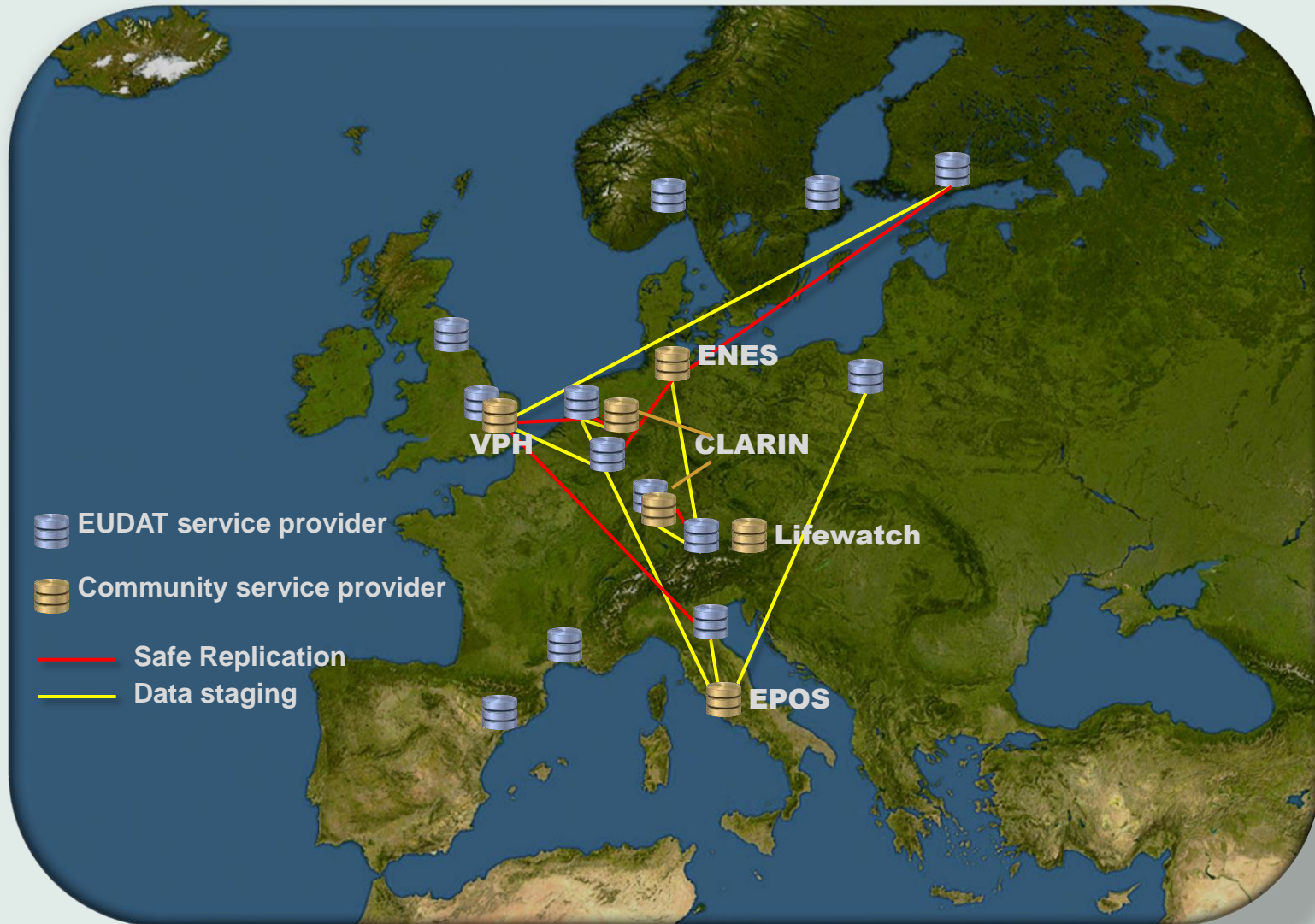




# EUDAT : une grille européenne de données

*Etat d'avancement du projet au 01/11/2012*

- Novembre 2012 : déploiement des pilotes



- Fin 2012 :
  - Pilotes du projet déployés – validation de l'architecture
  - Préparation des nœuds de la vague de déploiement n°2
- 2013 :
  - Déploiement des établissements de la vague n°2
  - Fin 2013 : CINES est un nœud à part entière de la grille
  - Préparation des nœuds de la vague de déploiement n°3
- 2014 :
  - Déploiement des établissements de la vague n°3

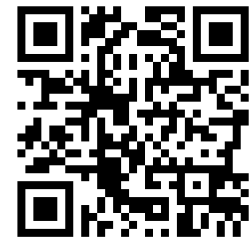


# Questions & Réponses



[marion.massol@cines.fr](mailto:marion.massol@cines.fr)

Pour plus d'information :



web

[www.cines.fr](http://www.cines.fr) (rubrique archivage)



## Annexes

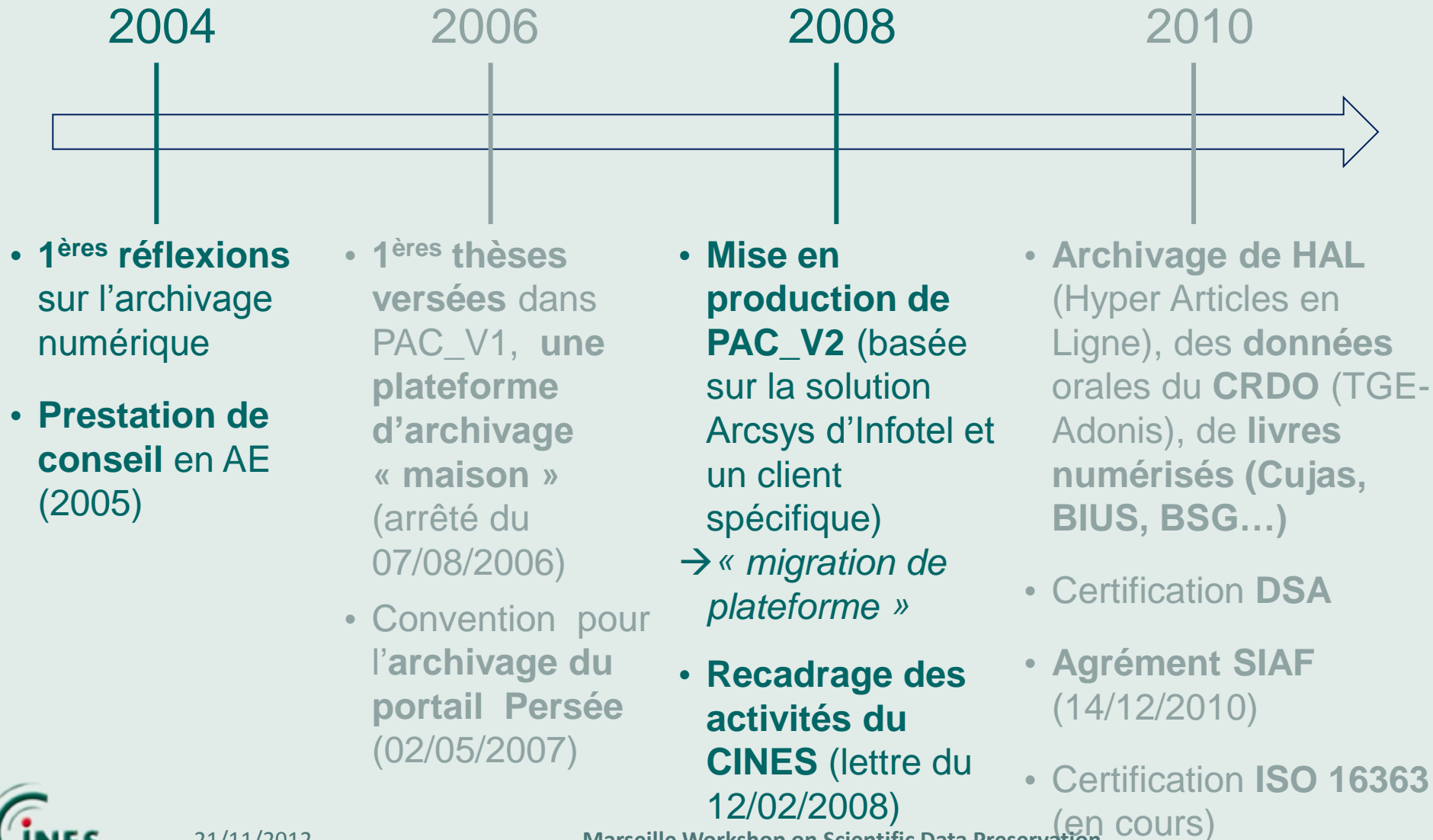
## Mise en place de procédures d'assurance qualité pour atténuer l'impact des quatre principaux risques identifiés lorsqu'ils se réalisent

Contrainte	Solutions
Connaissance du contenu	<ul style="list-style-type: none"><li>• Utilisation de métadonnées</li><li>• Identification unique et pérenne des documents archivés</li></ul>
Format de fichier inconnu	<ul style="list-style-type: none"><li>• Privilégier les formats durables</li><li>• Identification, validation des formats</li><li>• Migration logique (conversion de formats)</li></ul>
Support physique détérioré	<ul style="list-style-type: none"><li>• Gestion du vieillissement des médias</li><li>• Migration physique (changement de support)</li></ul>
Logiciel ou matériel de lecture disparu	<ul style="list-style-type: none"><li>• Veille technologique et anticipation</li></ul>

- **Différentes institutions interviennent au bout de la chaîne fonctionnelle d'archivage selon l'origine et la nature des documents électroniques**
  - Archives départementales (Universités)
  - Archives Nationales (EPA Education Nationale/Enseignement Supérieur)
  - BnF (Dépôt légal : publications, web)
  - INA (Dépôt légal : productions radiophoniques et télévisuelles)
  - CNC (Dépôt légal : production du cinéma et de l'image animée)
  - CINES (Thèses)
- **Les projets de numérisation/valorisation de documents papier échappent à ce contexte**
  - L'original papier sera conservé dans le service d'archives compétent
  - Possibilité de choisir l'institution archivant la version électronique du document

- **BnF – Bibliothèque Nationale de France**
  - Projet SPAR (Système de Préservation et d'Archivage Réparti)
- **CINES – Centre Informatique National de l'Enseignement Supérieur**
  - Projet PAC (Plateforme d'Archivage du CINES)
- **CNES – Centre National d'Etudes Spatiales**
  - Projet SIPAD (Système d'Information de Préservation et d'Accès aux Données)
- **SIAF – Service Interministériel des Archives de France**
  - Projet Pil@e (Pilote d'Archivage Electronique)
- **Ministère de l'Economie, de l'Industrie et de l'Emploi**
  - Projet ARCADE
- **Ministère des Finances (DGFIP)**
  - Projet ATLAS

# L'archivage au CINES en quelques dates...



- **OAIS - ISO 14721 : Reference model for an Open Archival Information System**
  - Modèle purement conceptuel, ne fait aucune recommandation technique
- **P2A Politique et pratiques d'archivage (sphère publique)**
  - Recommandations en termes d'architecture, moyens, sécurité, etc.
- **Standard d'échanges de données pour l'archivage électronique, versement, communication, élimination**
  - DAF, DGME, version 1.0, mars 2006.
- **Normes internationales de description archivistique**
  - ISAD-G – international standard for archival description, general
  - ISAAR-CPF – international standard archival authority record, corporate bodies, persons & families
- **Métadonnées descriptives de l'archive**
  - DCMI – Dublin Core Metadata Initiative
- **Identifiant unique et pérenne**
  - Interne, séquentiel, couplé à un identifiant persistant externe de type ARK
- **Empreintes numériques**
  - Hashing MD5, SHA-256

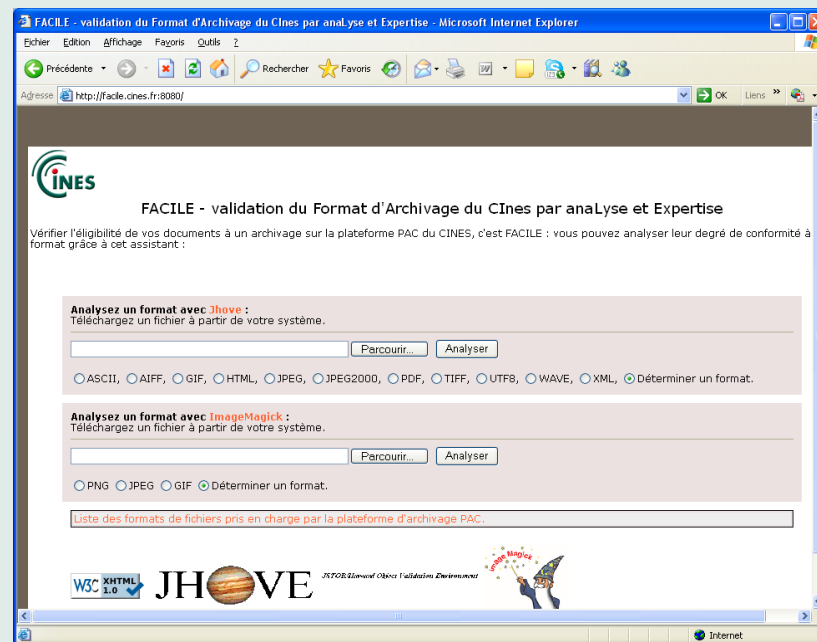


# Un outil pour contrôler la qualité des formats de fichier

## FACILE – validation du Format d'Archivage du Cines par anaLyse et Expertise

- Outil en ligne permettant de valider les fichiers par rapport aux spécifications de leur format ;
- Les contrôles effectués sont les mêmes que ceux effectués lors d'un dépôt de document ;
- Intègre les mêmes outils (Jhove, Imagemagick, DROID) que la plateforme d'archivage du CINES;
- Permet une validation des fichiers avant dépôt de la part du producteur.

<http://facile.cines.fr/>



## La Plateforme d'archivage pérenne au CINES – PAC v2.0

- Capable de gérer de larges volumes (2x40 To)
- En exploitation depuis Mai 2008.
- Volumétrie archivée : ~ 17To (au 1<sup>er</sup> Janvier 2012)



Tous les projets d'archives à long terme partagent la même plateforme

- Mutualisation de l'infrastructure matérielle d'archivage ;
- Protocole de versement générique ;
- Diminution des coûts de mise en place et d'exploitation.

## 1. Des projets en exploitation

- Archivage des thèses électroniques
  - Documents nativement au format électronique versés par l'ABES
- Archivage des revues SHS du portail Persée
  - Documents issus de la numérisation de revues dans le cadre du programme Persée
- Archivage des documents déposés dans les archives ouvertes
  - HAL – Hyper Article en Ligne du CCSD
- Archivage de documents sonores issus de la recherche dans le domaine de l'oral
  - Projet pilote CRDO dans le cadre du programme SHS du TGE-Adonis
- Archivage des ouvrages numérisés en histoire de la médecine de la BUPMC
- Archivage des ouvrages numérisés en histoire du droit de la bibliothèque CUJAS
- Archivage des ouvrages numérisés en histoire de la médecine de la BIUSanté
- Archivage des ouvrages numérisés de la bibliothèque Ste Geneviève

## 2. Plusieurs projets en cours de réalisation

- Archivage intermédiaire de données primaires scientifiques ISAAC
- Archivage de bases de données de l'INSERM, de données de la recherche (ATILF, IRHT...)

**La réussite au niveau national d'une stratégie pour l'archivage pérenne de documents électroniques produits par la communauté IST passe par la réalisation d'un certain nombre de défis – sur lesquels le CINES et la BnF sont actifs :**

- Acquisition d'une nouvelle compétence métier
- Sensibilisation de la communauté IST à la problématique de la préservation à long terme des documents numériques
- Sensibilisation des décideurs sur l'importance de l'enjeu
- Émergence et reconnaissance d'acteur(s) majeur(s) dans le domaine de l'archivage pérenne pour l'IST
- Intégration de la chaîne fonctionnelle de préservation au cycle de vie du document numérique : de la production à l'archivage

**Le CINES est maintenant un acteur reconnu du domaine de la préservation à long terme des documents numériques.**

- **Fort de la mission nationale que lui a confiée le MESR (centre officiel d'archivage des thèses électroniques)**
- **Il est appelé à jouer un rôle clé dans la réussite au niveau national d'une stratégie pour l'archivage pérenne des documents électroniques produits par la communauté Université-Recherche (Bibliothèque Scientifique Numérique).**

## Objectifs 2012-2013 :

- **La poursuite et le développement de tous ces projets,**
- **La consolidation et la professionnalisation du service**
  - Obtention de l'agrément SIAF pour la conservation d'archives publiques intermédiaires (10/2010)
  - Obtention de l'accréditation DSA – Data Seal of Approval (03/2011)
  - Démarche de certification du service d'archives en 2012 ou 2013 – ISO 16363

## Au niveau national :

- **Groupe de travail PIN (pérennisation de l'information numérique) au sein de l'association Aristote**
  - Lieu de rencontre et d'échanges entre informaticiens, archivistes et bibliothécaires, animé par le CINES, la BnF, le SIAF et le CNES
  - Réunions trimestrielles (environ 30 participants réguliers)
  - Un site web : <http://pin.association-aristote.fr/>
  - Une formation spécialisée (1 session par an)
- **Sous-groupe de travail « archivage pérenne » de la BSN (bibliothèque scientifique numérique) au sein du MESR**
  - Lieu de rencontre d'experts et d'acteurs de la conservation de documents numériques
  - Les recommandations du sous-groupe permettront au Comité de pilotage de définir les modalités d'organisation et de gouvernance de la BSN et de ses opérateurs, des indicateurs d'activité et des critères d'évaluation de l'activité.

## Au niveau international :

- **wePreserve** (<http://www.wepreserve.eu/>)
  - Synergie de plusieurs projets européens (DPE, CASPAR, PLANETS)
- **Alliance for Permanent Access** (<http://www.alliancepermanentaccess.org/>)
  - Coordination de la mise en place d'une infrastructure européenne pour l'archivage pérenne des données scientifiques
  - Groupe de travail pour le test de la certification de systèmes d'archivage électronique (ISO 16363)
- **Le consortium IIPC** ([www.netpreserve.org](http://www.netpreserve.org))
  - Acquisition, préservation et accès à l'information venant de l'Internet pour les générations futures
- **La conférence iPRES** (<http://www.slis.tsukuba.ac.jp/ipres2011/>)
- **De nouvelles communautés ?**

