

Préservation des Données à Long Terme & Open Data à l'IN2P3

Yonny CARDENAS

Journées Informatique IN2P3-IRFU

Port-Bail, 1-4 Octobre 2018

- Open Data / Ouverture de Données

Principe, politique, selon lequel les données "publiques" doivent être disponibles pour l'accès et la réutilisation par les citoyens et organisations.

Donnée numérique dont l'accès et l'usage sont laissés libres aux usagers.

Donnée diffusée de manière structurée selon une méthode et une licence ouverte garantissant son libre accès et sa réutilisation par tous, sans restriction technique, juridique ou financière.

- Open Data / Ouverture de Données
 - *principe*
 - *donnée numérique*
 - *accès et l'usage libres*
 - *réutilisation*
 - *diffusée de manière structurée*

- *information est un bien commun*

- Préservation des Données à Long Terme – Archive

Une Archive est une organisation chargée de conserver l'information pour permettre à une communauté d'utilisateurs cible d'y accéder et de l'utiliser.

Préservation numérique consiste en un certain nombre de méthodes qui visent à assurer l'accessibilité permanente aux matériaux numériques.

Un système pour garantir la préservation des données numériques: à faire en sorte que l'information reste lisible, compréhensible et réutilisable sur le long terme.

- Préservation des Données à Long Terme – Archive
 - *Information*
 - *conservation*
 - *accessible*
 - *utilisable et réutilisable*
 - *long terme*

 - *valeur ou patrimoine*

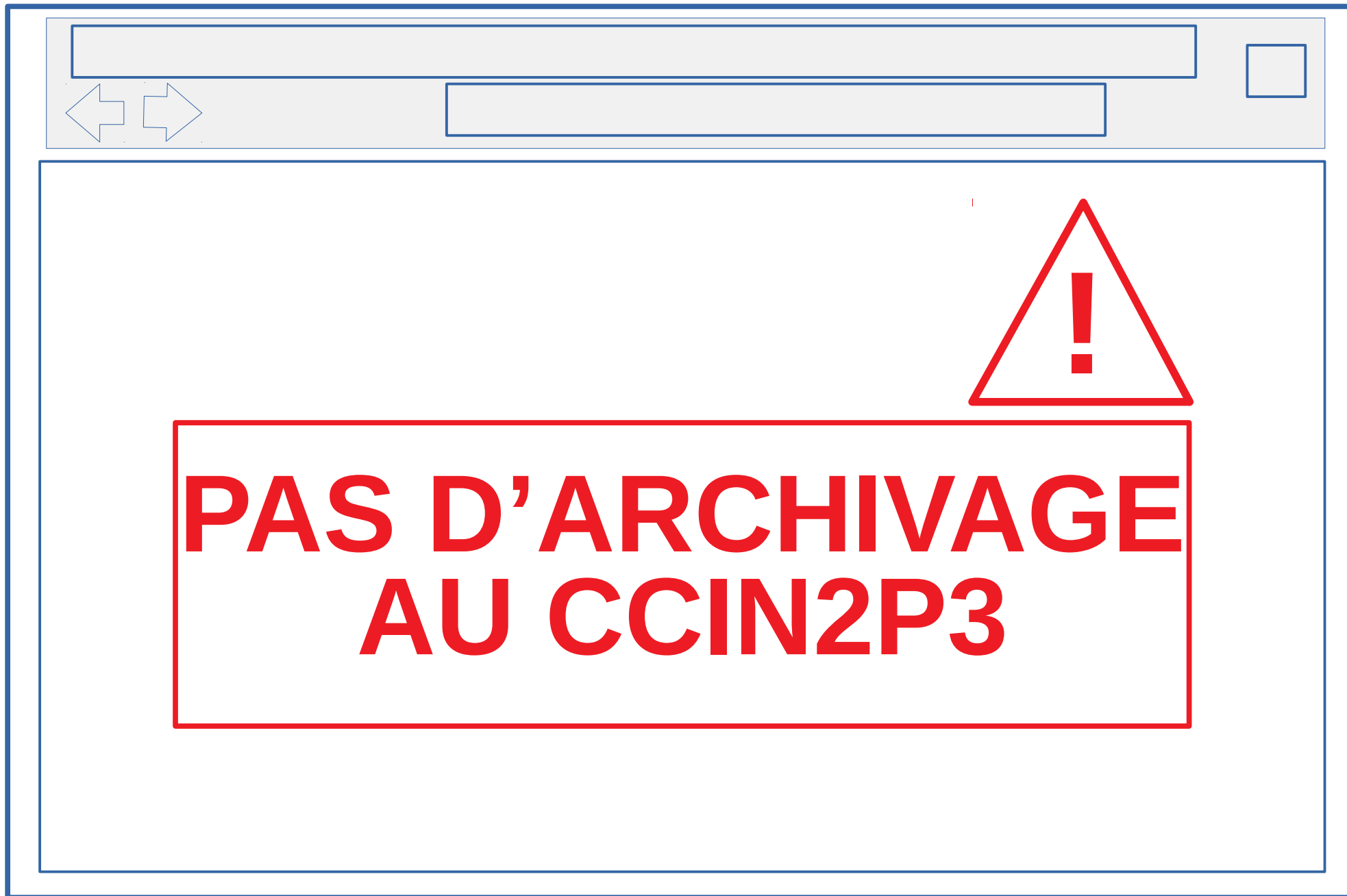
Motivations pour l'archivage

- Potentiel scientifique (propriétaire)
- Politique institutionnel (patrimoine)
- Pratique communauté scientifique
- Engagement (organisme de financement)
- Législation

Ne pas confondre avec :

- Copie
- Sauvegarde
- Réplication
- Stockage
- Migration de supports
- Gestion de versions
- Stockage grille
- Stockage cloud





Attention aux engagements !!!

- Engagements de la part de :
 - Laboratoire
 - Projet de recherche
 - Chercheurs, ingénieurs, ...
 - ...
- Engagements vis-à-vis des des agences et organismes de financement
- Car au CC-IN2P3 :
 - Aucune démarche de préservation/archivage
 - Aucune certification
 - Aucune démarche audit (interne ou externe)
 - Pas de politique institutionnelle d'archivage à long terme

Problématique :

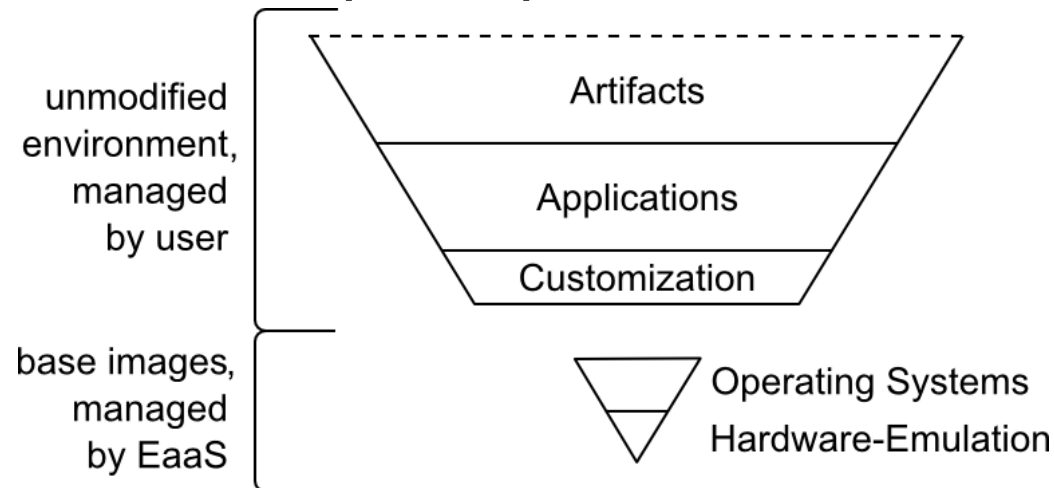
Grande vulnérabilité données numériques par rapport au temps

- Supports d'enregistrement
 - conservation (CD, DVD, Tapes, Disques, ...)
 - dégradation
 - obsolescence technologique
- Formats de représentation
 - obsolescence formats
(e.g. textes, images, audiovisuels,...)

Pérennisation – Contexte

Stratégies de pérennisation de l'information

- Migration
 - Supports
 - Formats
- Émulation
 - Approche fonctionnel
 - Objets numériques complexes
 - Émulation as a Service (EaaS)

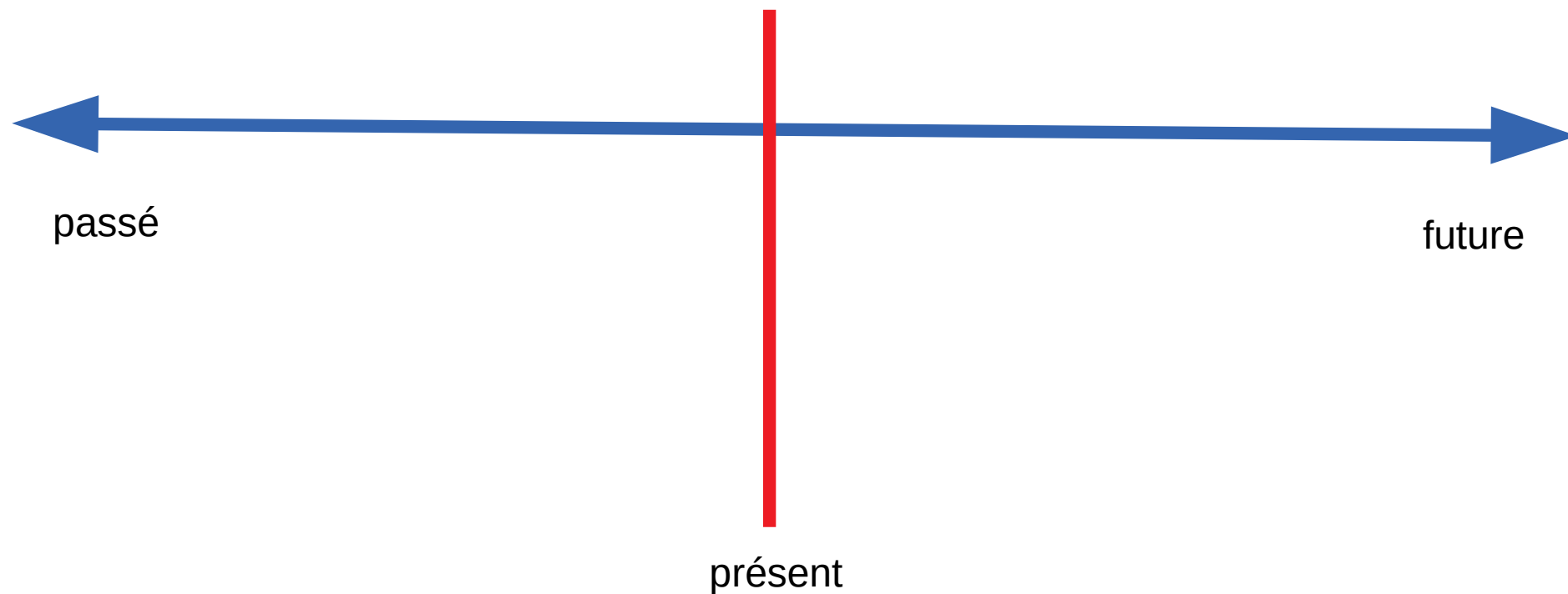


crédits : <http://eaas.uni-freiburg.de>

Stratégies de pérennisation de l'information

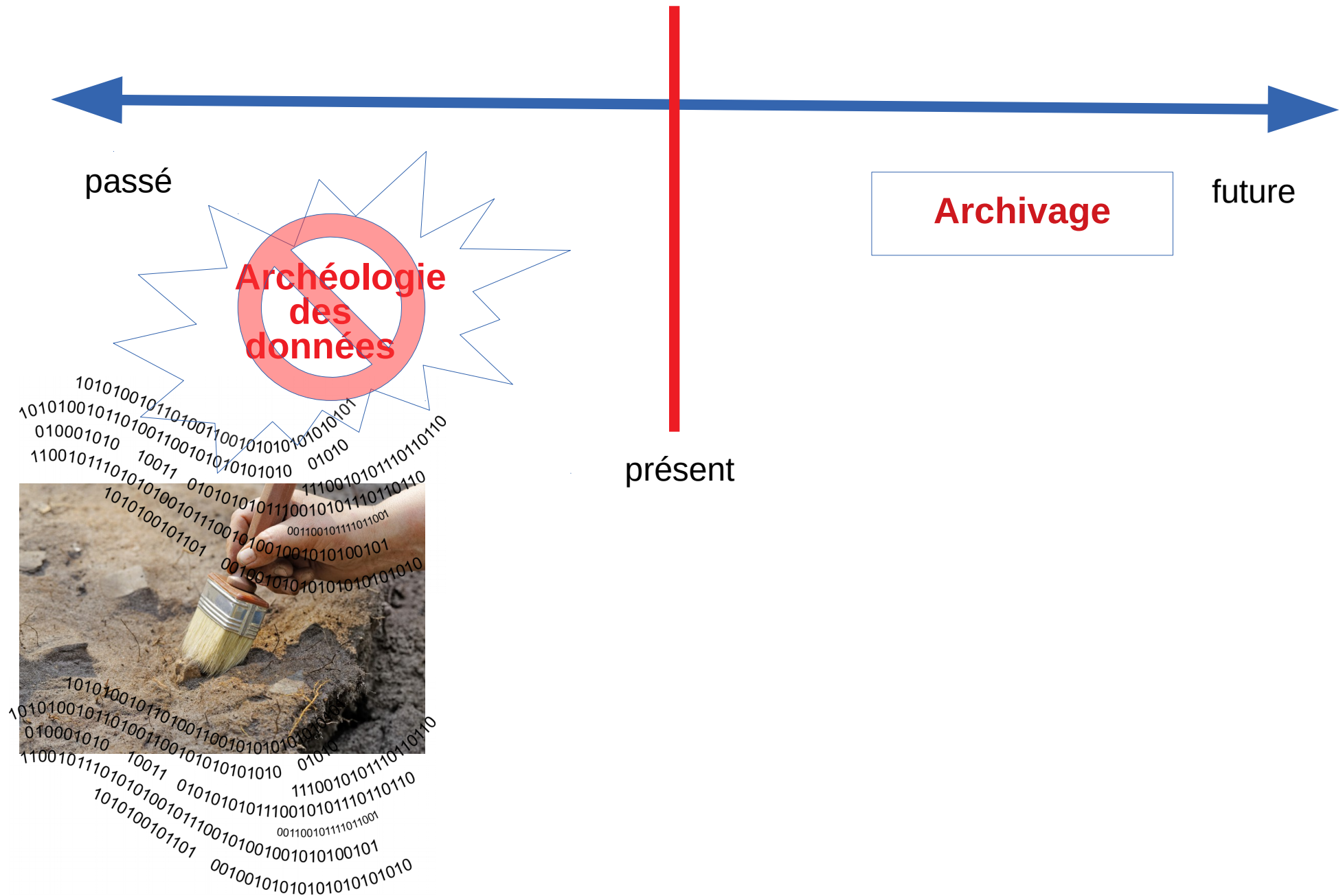
- Organisation et documentation
 - Normes et Standards
 - Métadonnées
 - Descriptives
 - Structure
 - Techniques
 - Préservation
 - Droits
- Gestion de risques
- Gestion de coûts
- Processus d'amélioration continue

Place de l'Archivage



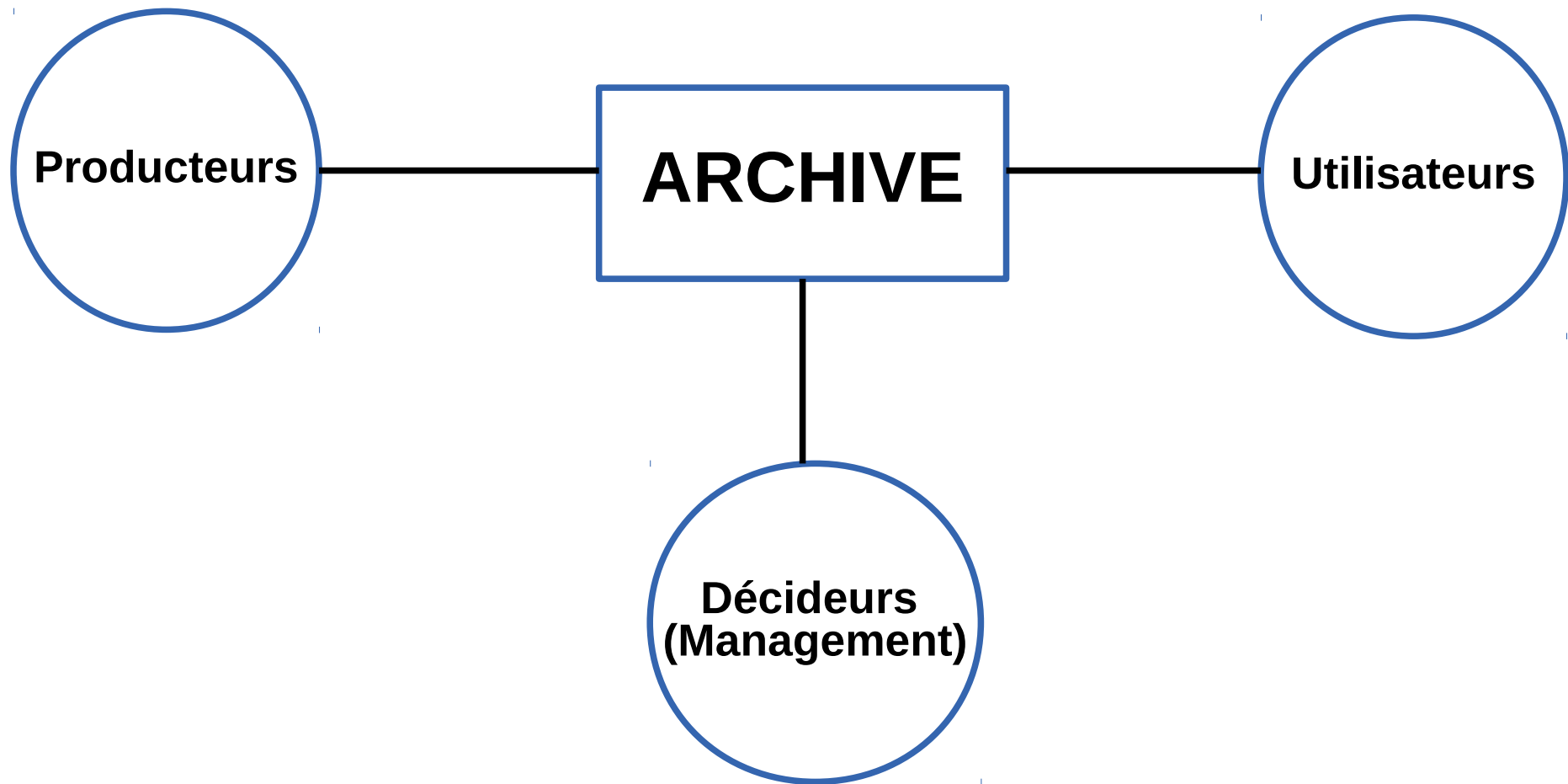
DONNÉES

Place de l'Archivage

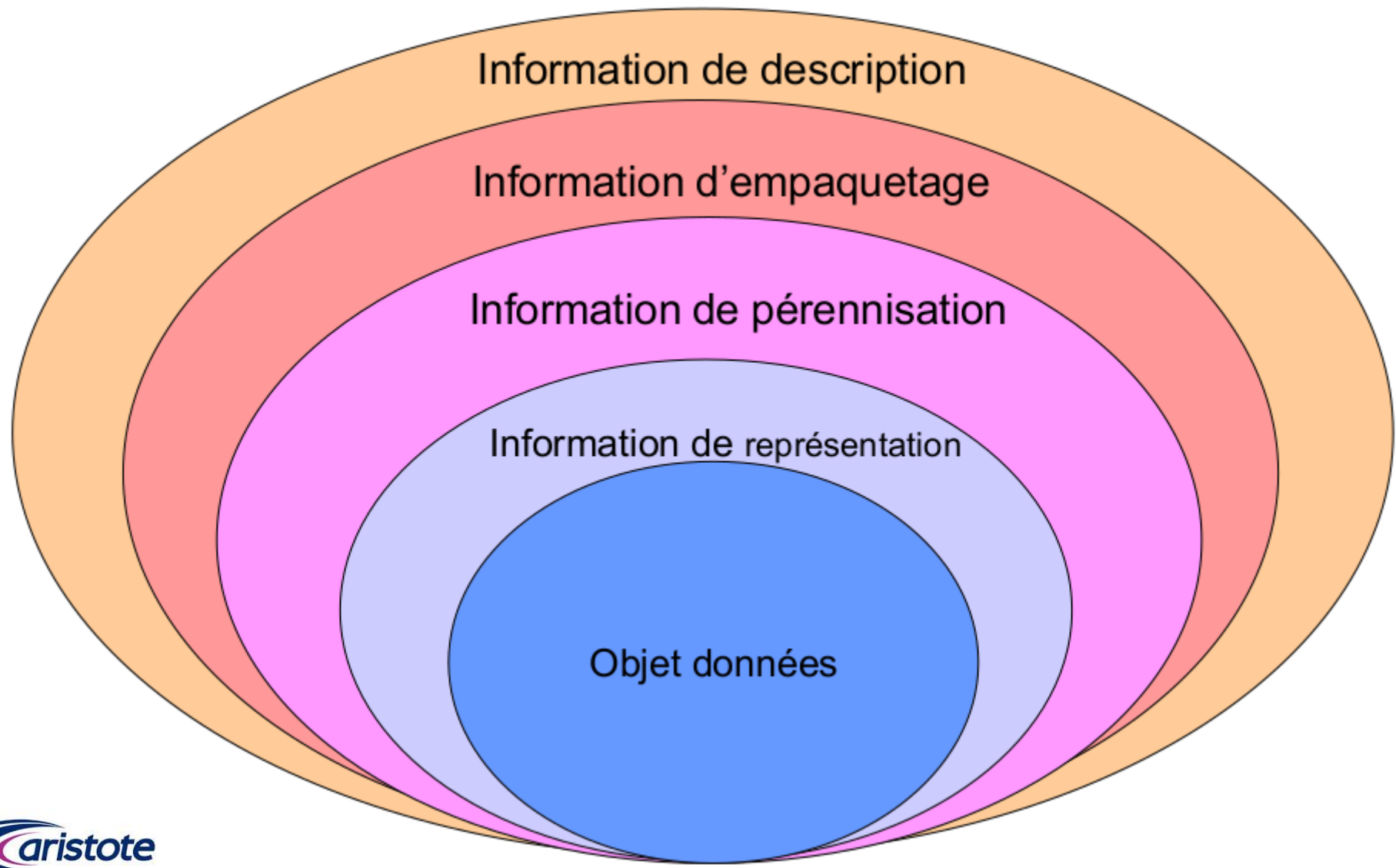


OAIS

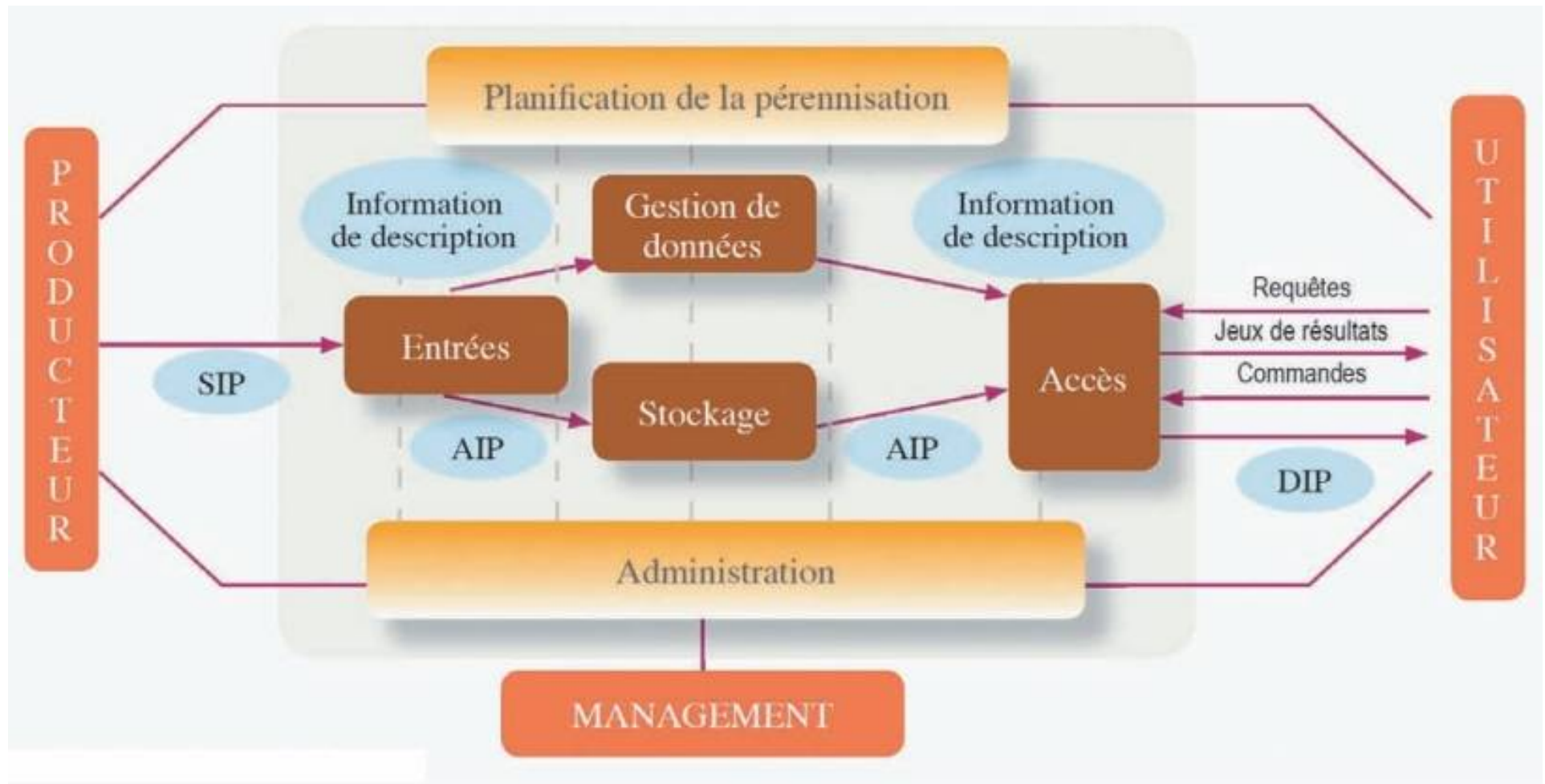
Systeme Ouvert d'Archivage d'Information



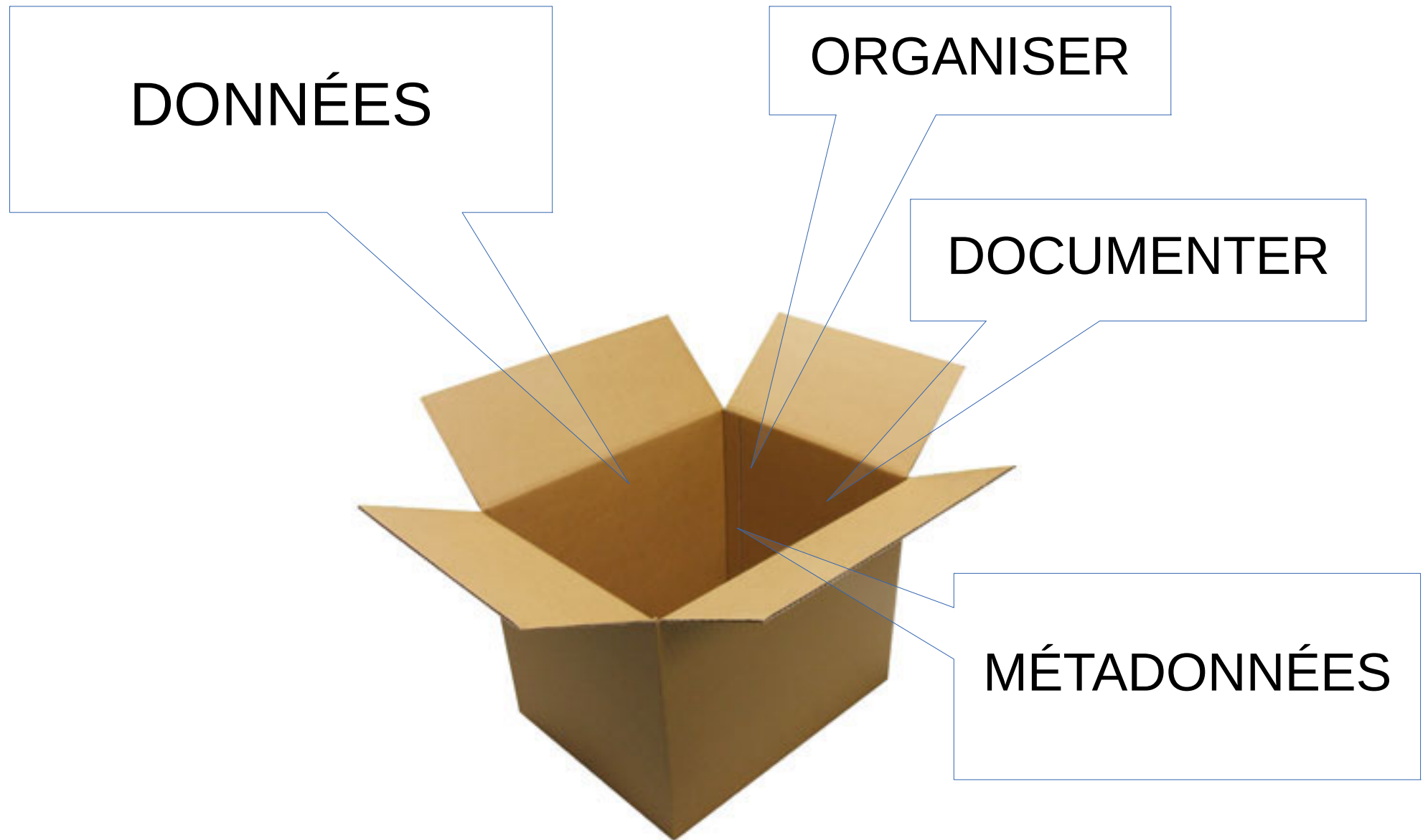
OAIS : Modèle d'Information



OAIS: modèle fonctionnel



OAIS: Paquet d'Information





Préservation & Découverte



Objective : Réutilisation de l'information

Quelques processus en commun à mutualiser :

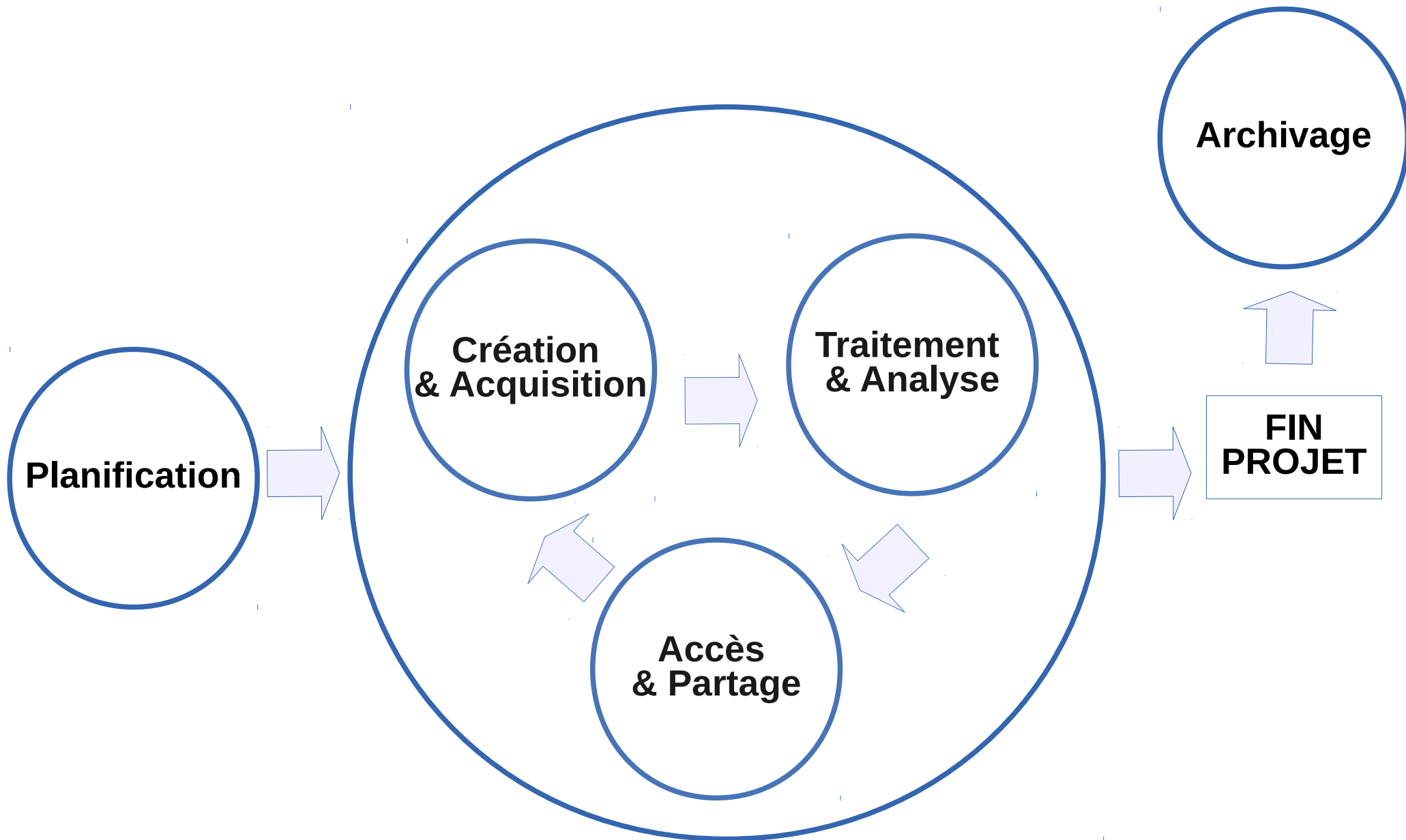
- Organisation et structuration de données
- Documentation
- Description (Métadonnées)
- Planification (Plan de Gestion des Données)

Préservation

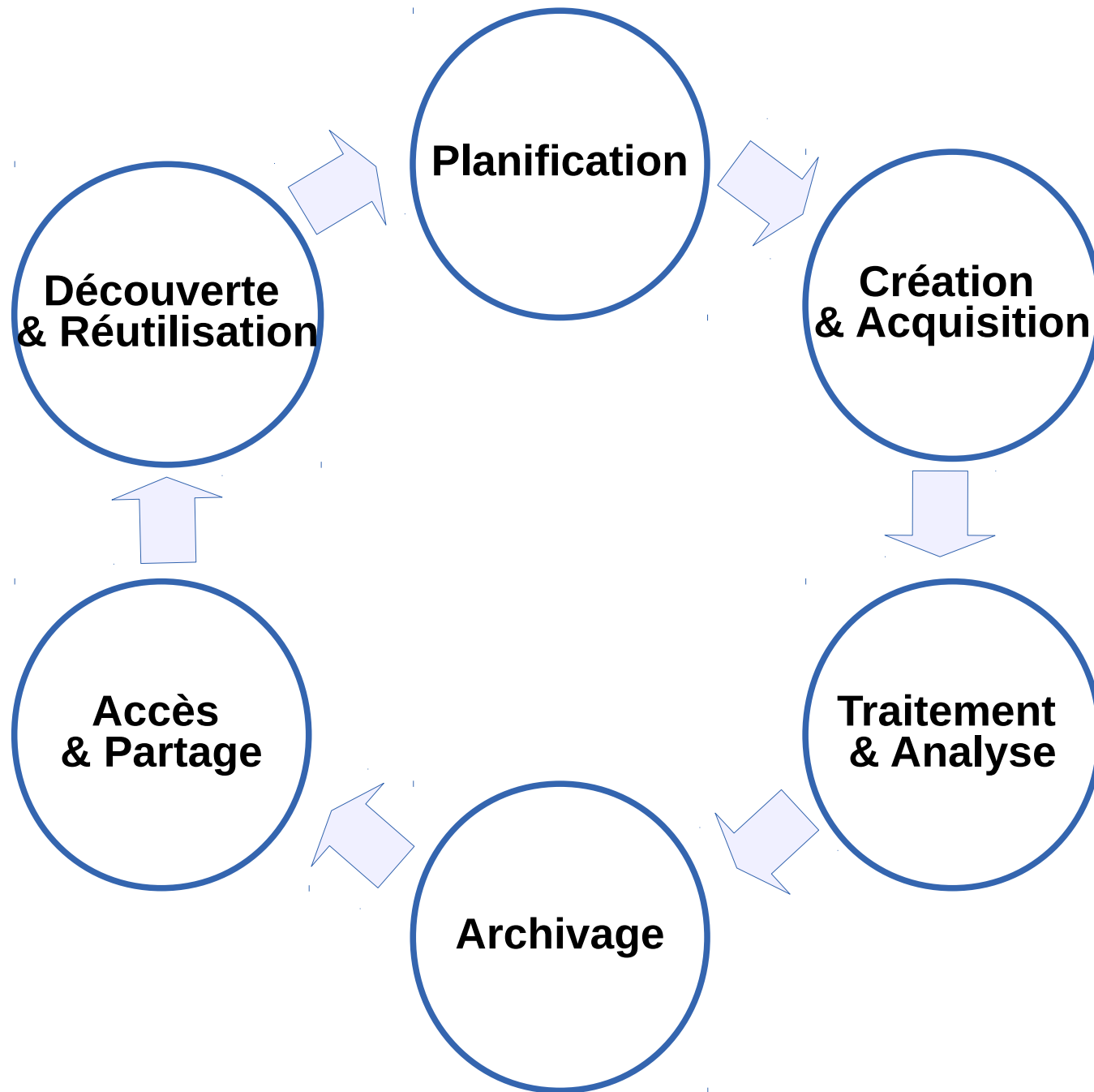


Open Data

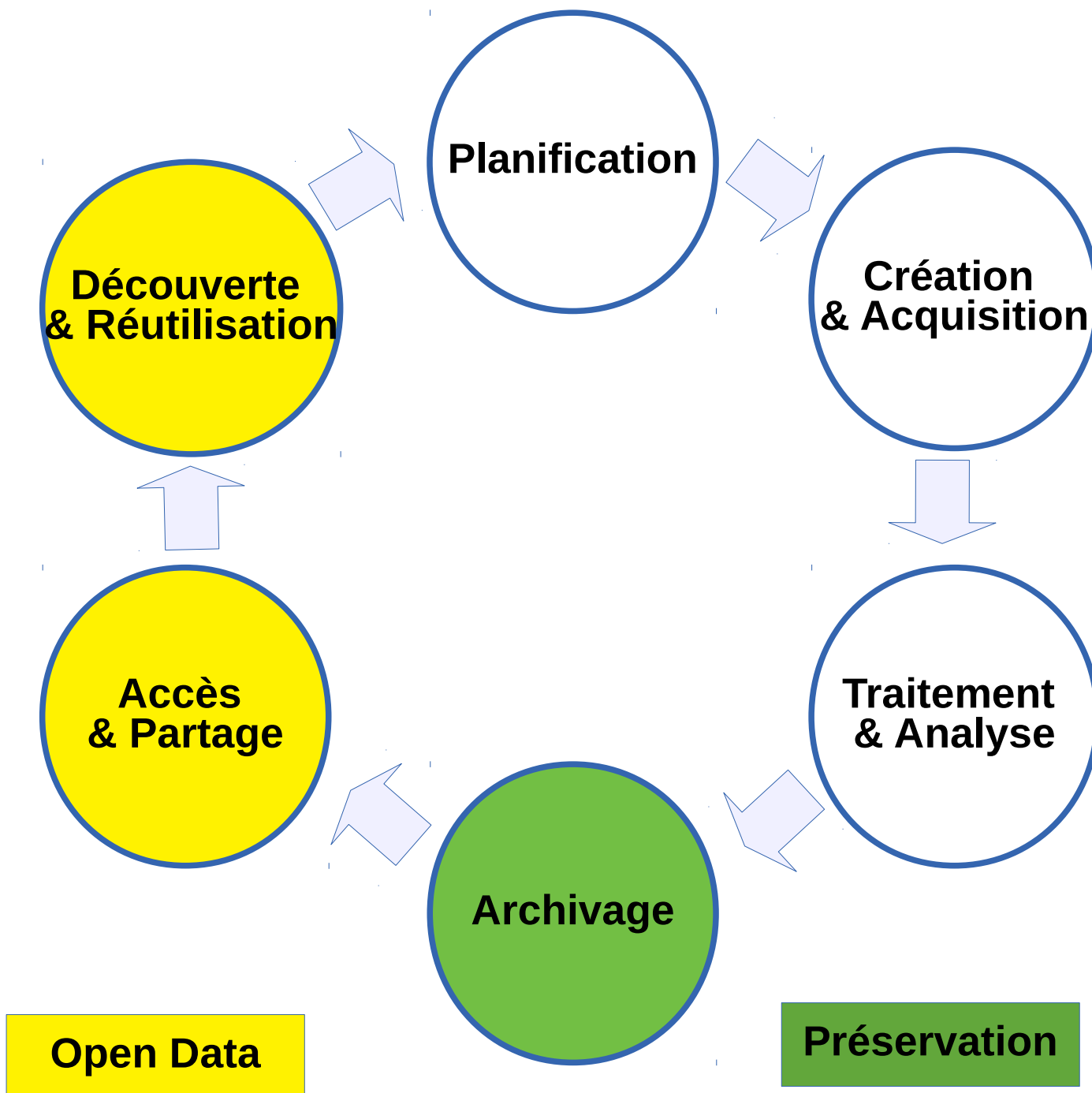
Cycle de vie des données de recherche



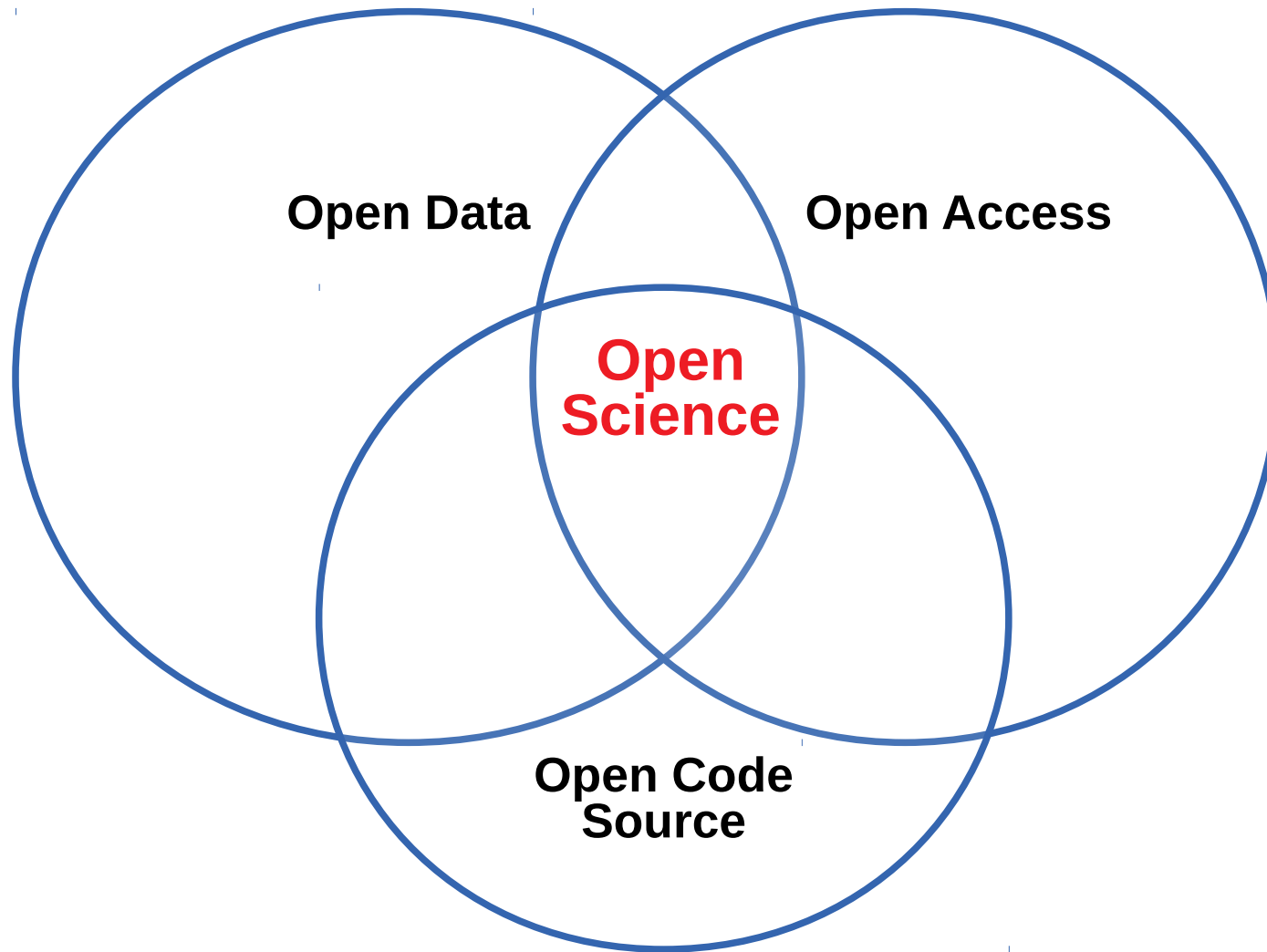
Cycle de vie des données de recherche



Cycle de vie des données de recherche



Contexte Open Science



- Pas d'archivage au CC-IN2P3
- Données scientifiques brutes (raw data)
- Volumétrie considérable (téraoctets ou plus)
- Diversité de formats et utilisations
 - Stratégie de migration de format n'est pas viable
- Travail porté sur les information de description, gestion global de données.

- Nécessité de la mise en œuvre d'une plate-forme d'archivage interne?
- Une plate-forme externe?

Dans tous le cas, nous avons besoin de:

- Stratégie institutionnel de préservation des données
- Élaboration de charte de préservation
 - responsabilités
 - obligations
 - parties
 - ...
- Définition de fonctionnalités et disponibilité du service
- Définition de politique de sécurité
- Allocation des moyens

Gestion de Données de Recherche

- Plan de Gestion des Données
- Catégorisation des jeux de données à l'aide de métadonnées ;
- Définition des différents workflows de données (physiciens GANIL, Collaboration,...) ;
- Stockage local et **archivage** au Centre de Calcul ;
- Mise en place de DOI sur les jeux de données ;
- Partage des données (**Open Access**)
- Création d'un portail des données ;
- Utilisation de licences sur les données ;

(crédits : Benjamin Fusberti)

Remerciements

- BnF - Bibliothèque Nationale de France
- CINES
- Le groupe de travail PIN
(pérennisation des informations numériques)
association Aristote
- INIST/CNRS
- Réseau Thématique Démocrite / IN2P3

Préservation des Données à Long Terme & Open Data à l'IN2P3

Yonny CARDENAS

Journées Informatique IN2P3-IRFU

Port-Bail, 1-4 Octobre 2018



Documentaire

Réalisateur : Vincent Amouroux
Producteurs : ARTE France, ZED
52 min.
Diffusion : 2015

Récompenses:
Festival Pariscience 2015
(Prix Spécial du Jury)

Version VOD (boutique.arte.tv)

arte