

EUDAT and CINES data preservation services

Workshop PREDON

Stéphane Coutin – 05 nov 2014



Centre Informatique National de l'Enseignement Supérieur

- Basé à Montpellier (Hérault, France)
- EPA créé en 1999, succédant au CNUSC (Centre National Universitaire Sud de Calcul) – créé en 1980
- Placé sous la tutelle de la DGRI (Direction Générale de la Recherche et de l'Innovation) et de la DGESIP (Direction Générale pour l'Enseignement Supérieur et l'Insertion Professionnelle) du Ministère de l'Enseignement Supérieur et de la Recherche
- Les missions :
 - Calcul numérique intensif
 - Archivage pérenne de documents électroniques
 - Activité transversale : hébergement d'environnements informatiques
- Plus d'information : <http://www.cines.fr/>



- **Service definition (ITIL v3) :** *“A means of delivering value to Customers by facilitating Outcomes Customers want to achieve without the ownership of specific Costs and Risks.”*
- **Agenda:**
 - EUDAT projects and services building
 - Using services in projects
 - CINES service catalog

EUDAT project and services building

European Data



EUDAT

- Start date: 1st October 2011
- Duration: 36 Months
- Budget: 16.3 M€ (9.3M€ EC)
- EC Call: INFRA-2011-1.2.2
- Consortium: 25 partners from 13 countries
 - National data centers, technology providers, research
- Objectives:
 - Cost-efficient and high-quality CDI
 - Meetings users' needs in flexible and sustainable way
 - Across geographical and disciplinary boundaries



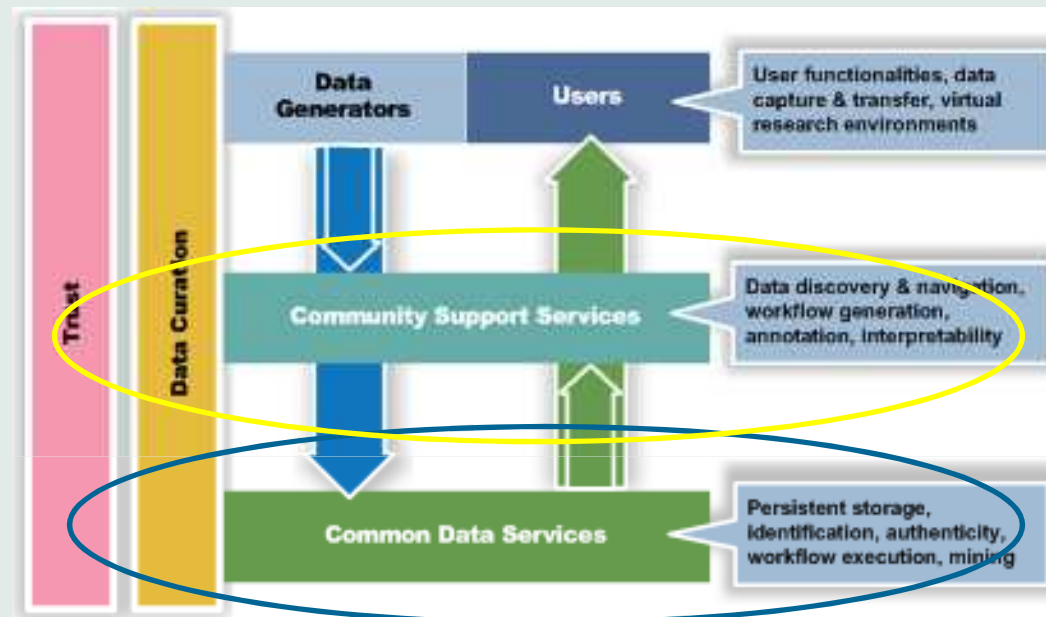
<http://www.eudat.eu>



Max Planck Institute for Psycholinguistics



Towards a Collaborative Data Infrastructure

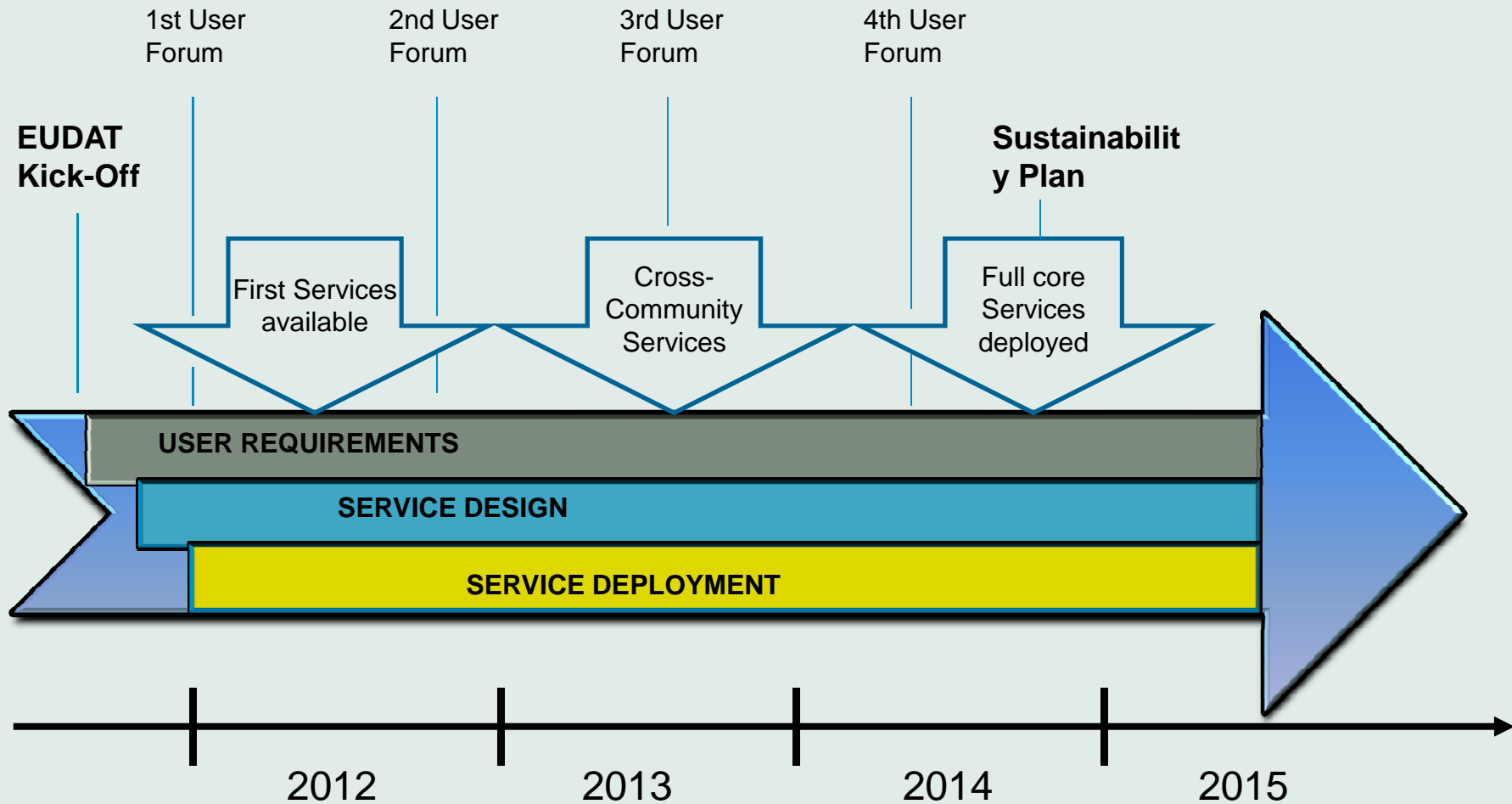


- EUDAT will focus on building this generic data infrastructure layer and offer a trusted domain for long term data preservation accompanied with related services to store, identify, authenticate and mine these data.
- This need be done in close collaboration with the Communities
 - Core services must match the requirements of the communities
 - Community services can also be incorporated into the common data service infrastructure when they are of use to other communities.

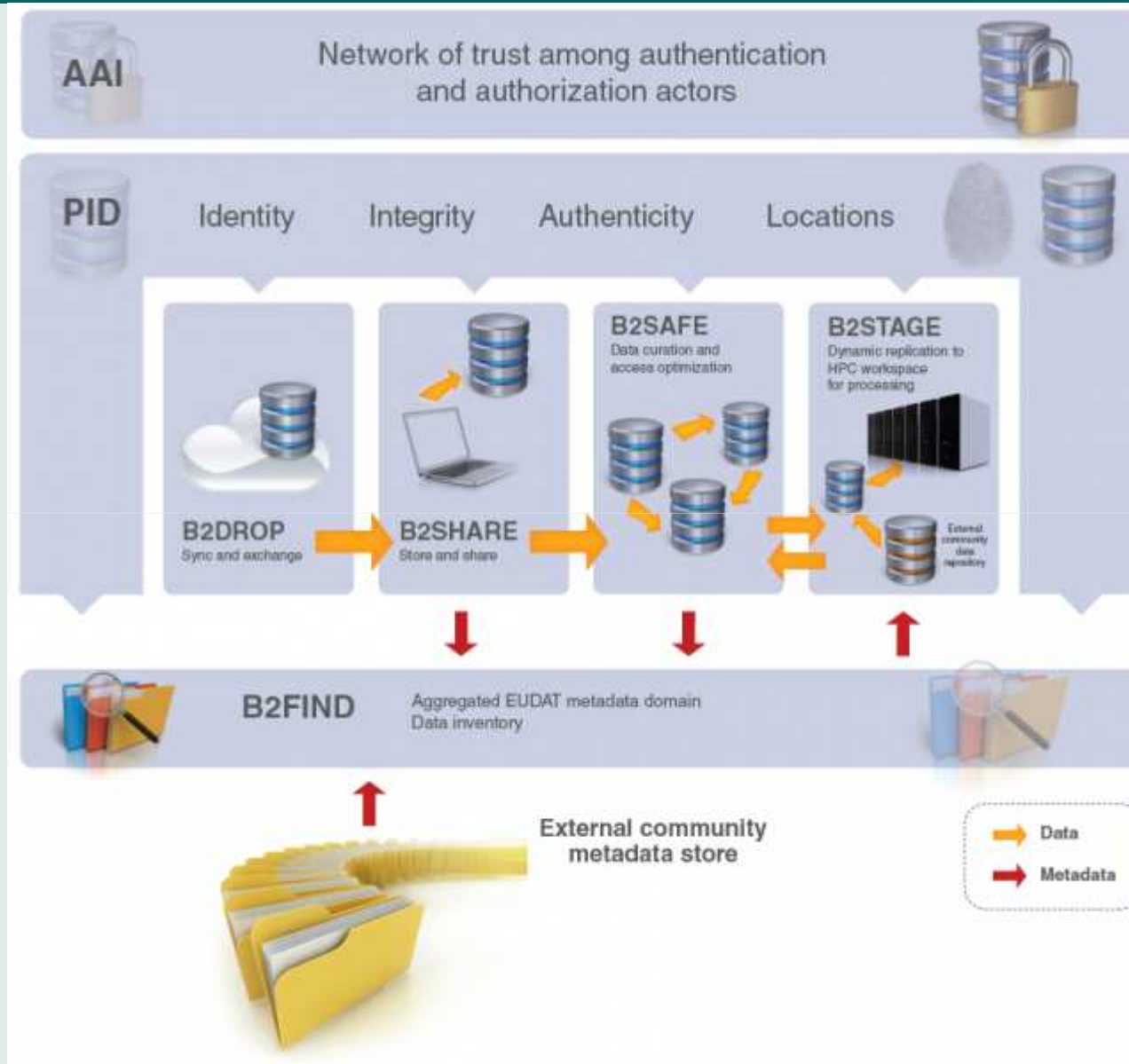
Communities ↔ Data Centers



EUDAT planned timeline



EUDAT services today



B2SAFE Phase 1 - Operational prototypes

- **Objectives :**

- Build quickly (agility)
- Experiment technology and features

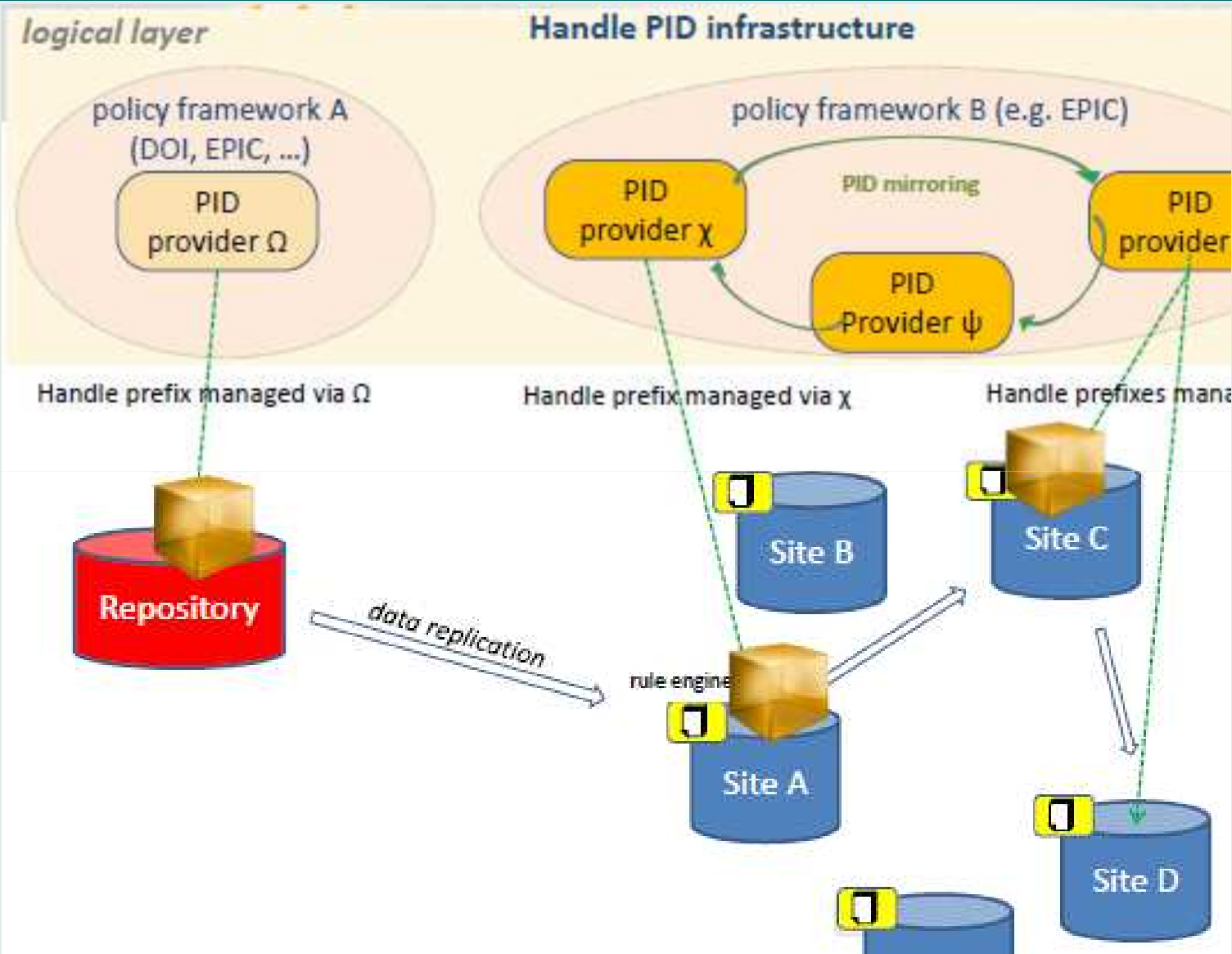
- **Some initial common building blocks:**

- High level requirements
- PID service
- Central tools (helpdesk, monitoring, resource coordination)

B2SAFE Phase 2 - Standard scalable solutions

- **Cross community features delivery**
- **Roadmap and versioning**
- **Components integration**
- **Simplify deployment**

B2SAFE core service



Using services in projects

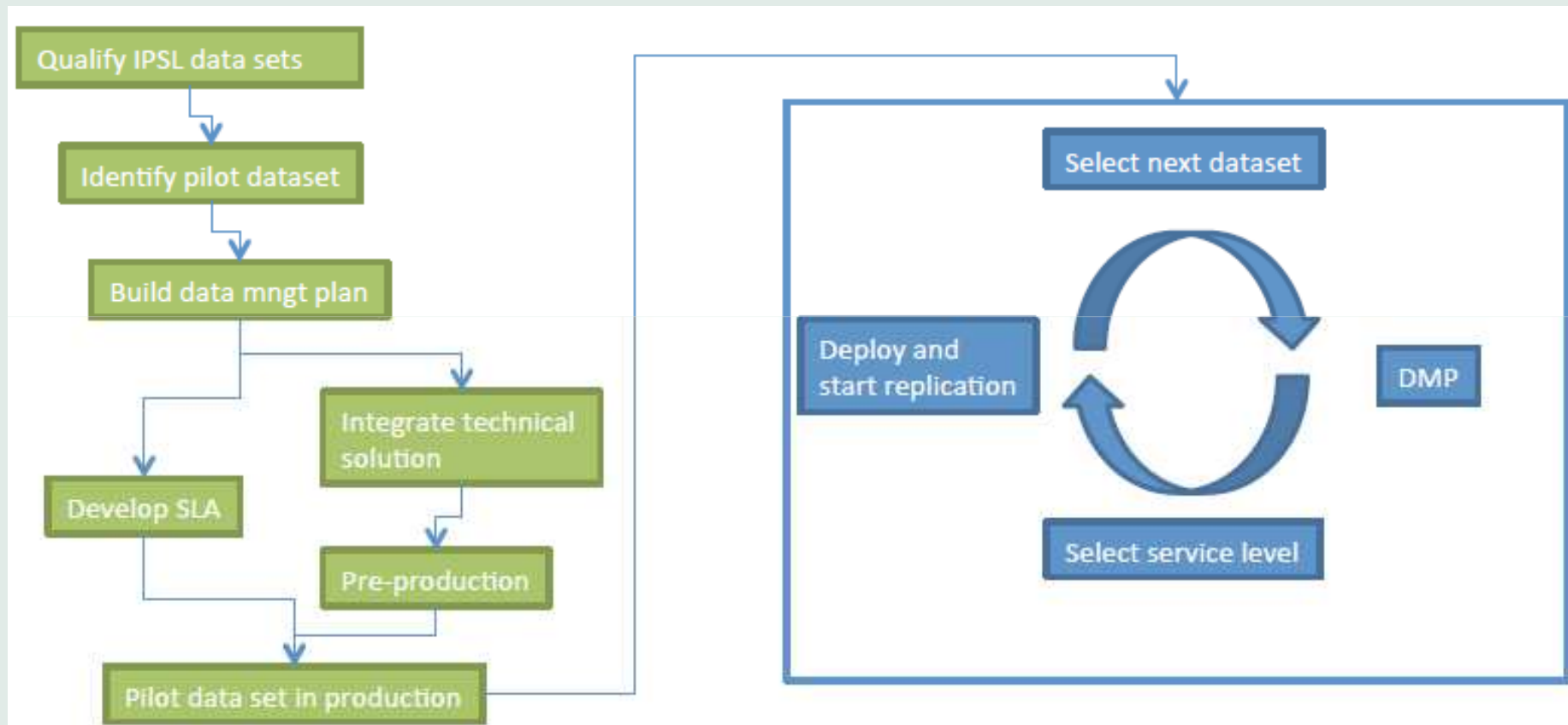
Project example : IPSL (Paris)

- **Applied to call for collaboration (Jan 2014)**
- **Environmental data**
- **Multiple datasets (some being non reproducible)**
- **More than 500 Tb stored on file system on 2 sites**
- **Initial usage : Avoid data loss (backup / restore)**
- **Potential usage : Reference data, Bring close to computation.**

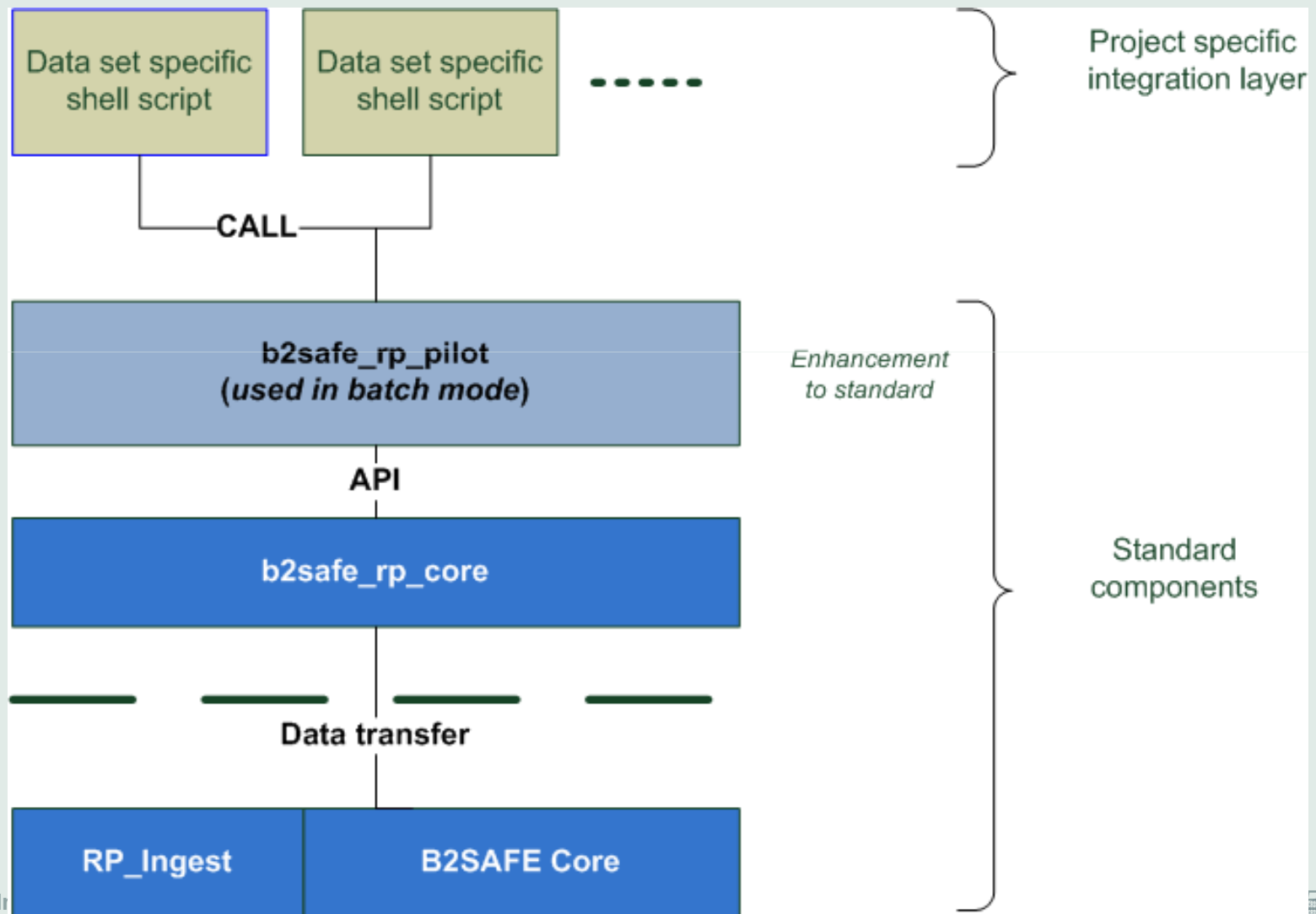
IPSL replication : an integration project

- **Clear direction : no specific solution development unless it cannot match the standard solution roadmap**
- **This will guarantee :**
 - Simpler support (shared with the consortium)
 - Ability to benefit from enhancements without effort

Project life cycle



IPSL replication : an integration project



CINES service catalog

Les services d'archivages du CINES

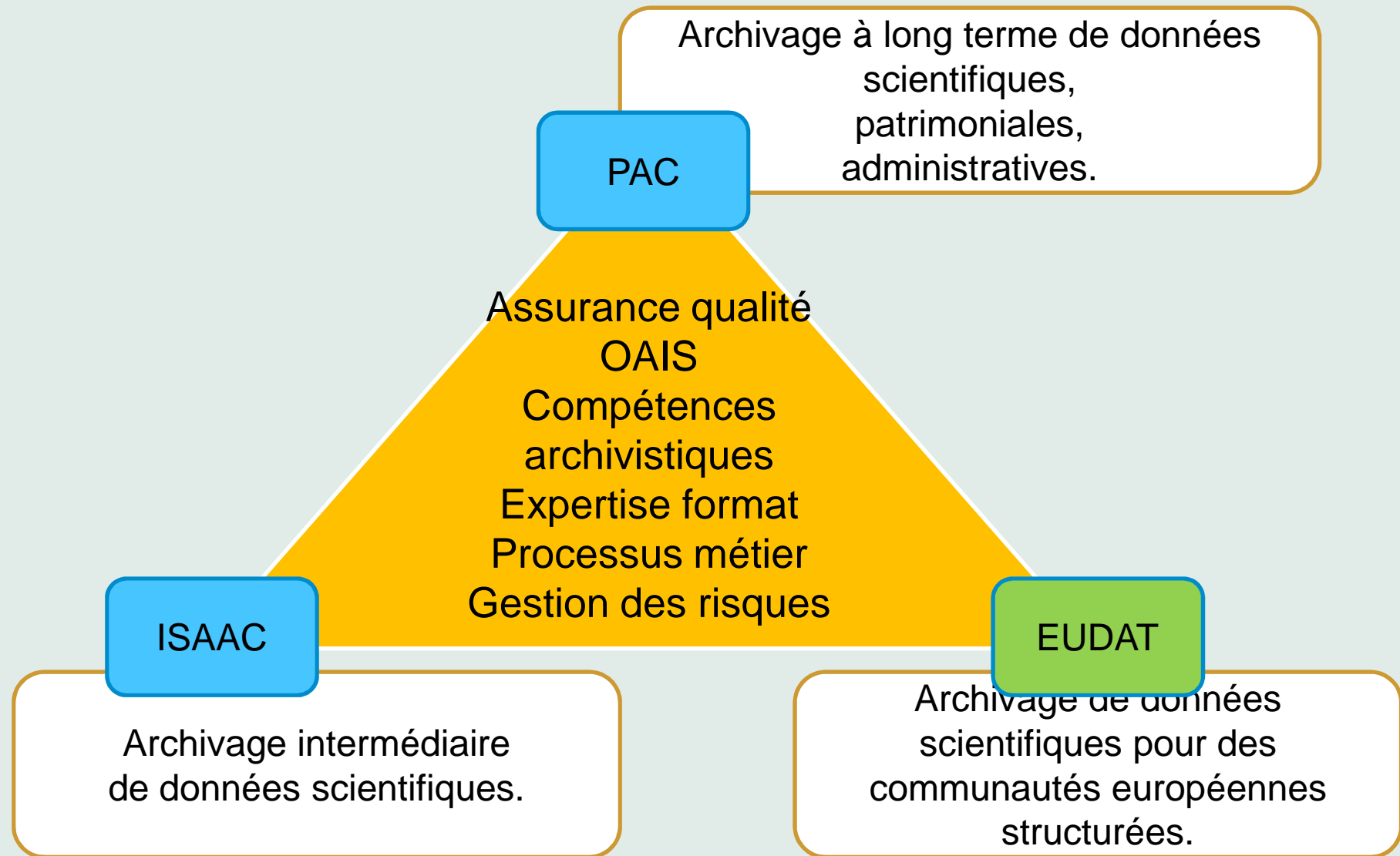
PAC

- **archivage intermédiaire et à long terme de données administratives, patrimoniales et scientifiques**
 - Mandat pour l'archivage des thèses électroniques soutenues en France (arrêté du 7 août 2006)
 - Agréé pour l'archivage intermédiaire par le SIAF
 - Périmètre opérationnel : données de l'enseignement supérieur et de la recherche
 - Partenariat avec le TGE Adonis : archivage et diffusion des données numériques en SHS

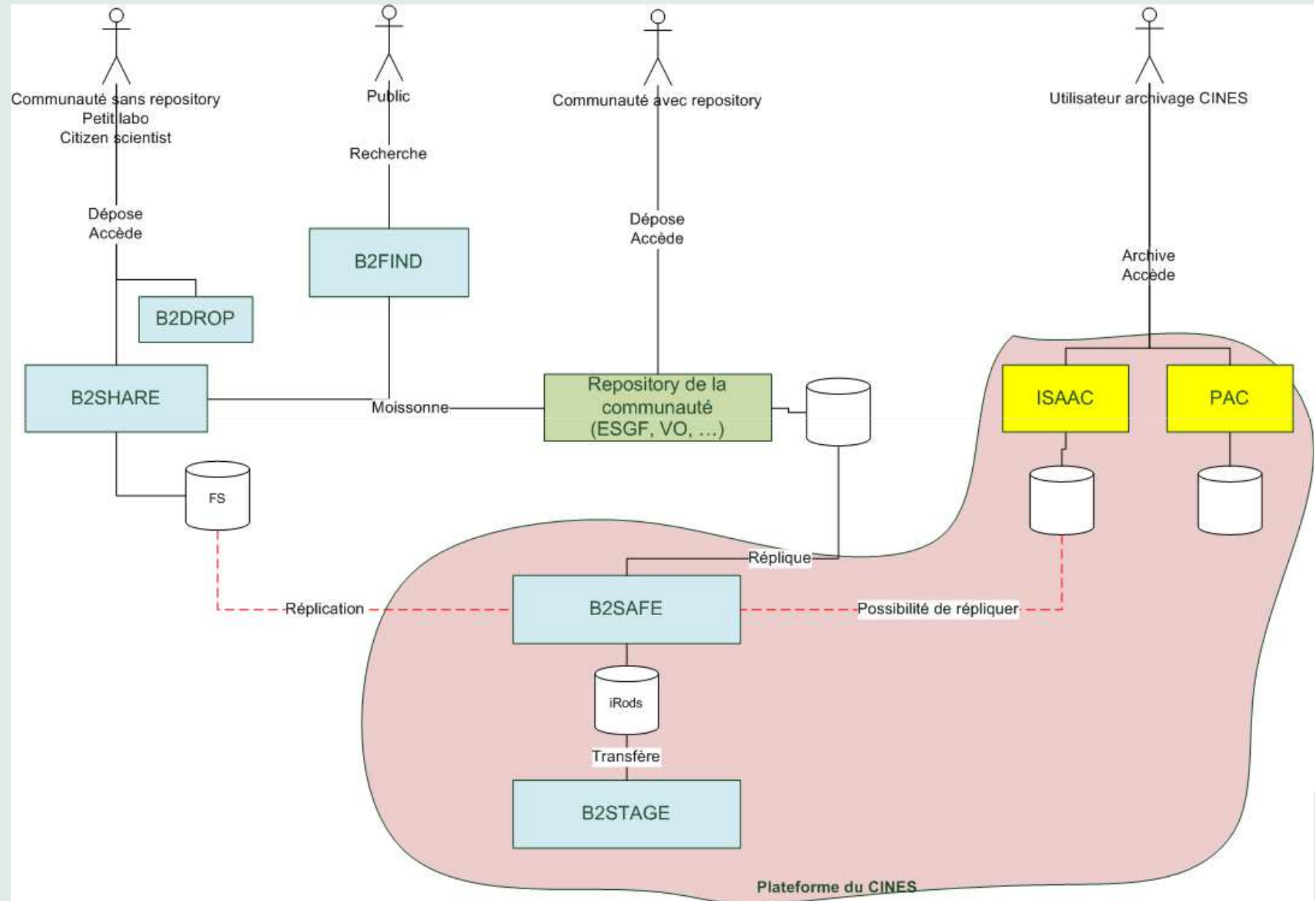
ISAAC

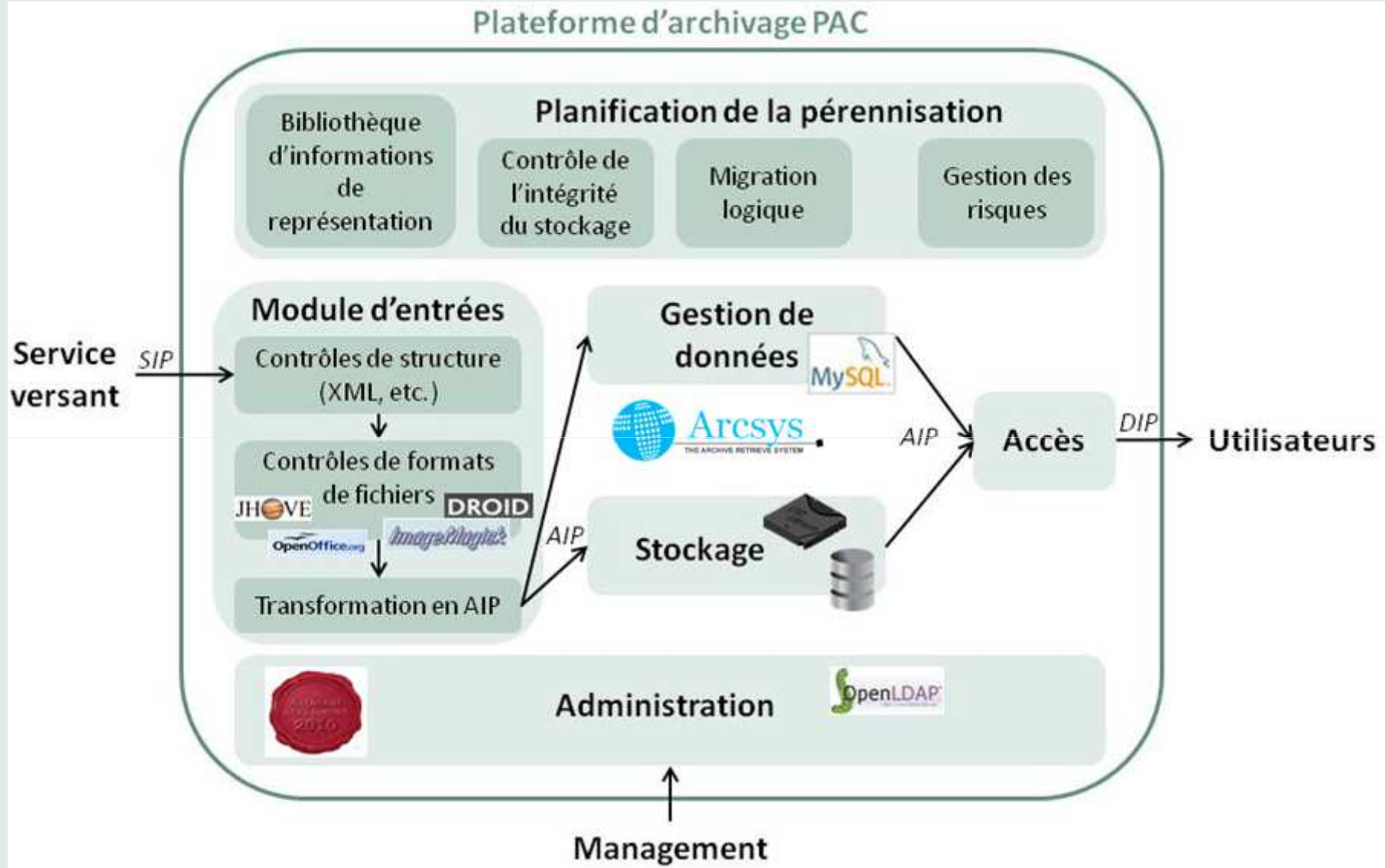
- **archivage intermédiaire de données scientifiques**
 - Dimensionné pour des petites structures ayant de grands volumes de données
 - Une donnée organisée et validée par des communautés d'experts
 - Un travail scientifique valorisé par le partage et la diffusion

Des niveaux de service adaptés aux besoins



Architecture fonctionnelle





Questions & Réponses

