

# Long Term Data Preservation

Marcello Maggi  
INFN/Bari

# Definizione

Data Preservation:

Meccanismo di conservazione della capacità di utilizzare dati digitali molto oltre il limite di esistenza dell'organizzazione che ha prodotto tali dati ovviando all'obsolescenza digitale

Include il Bit Preservation (ma molto altro)

Esclude la pronta disponibilità

# Storia

Agenzie Spaziali molto attive da decenni  
Con Standards e Open Access ai dati

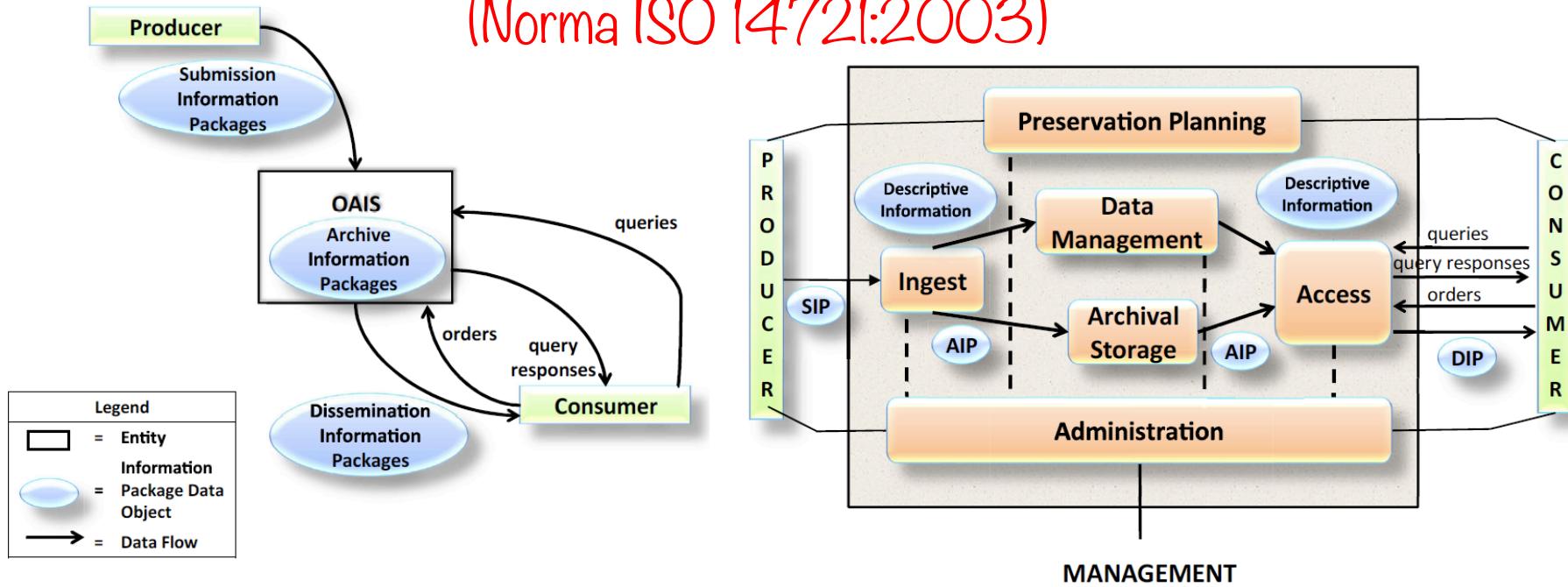
- FITS (NASA)
- OAIS (ESA)

Esigenza per Osservazioni che per definizione  
sono uniche

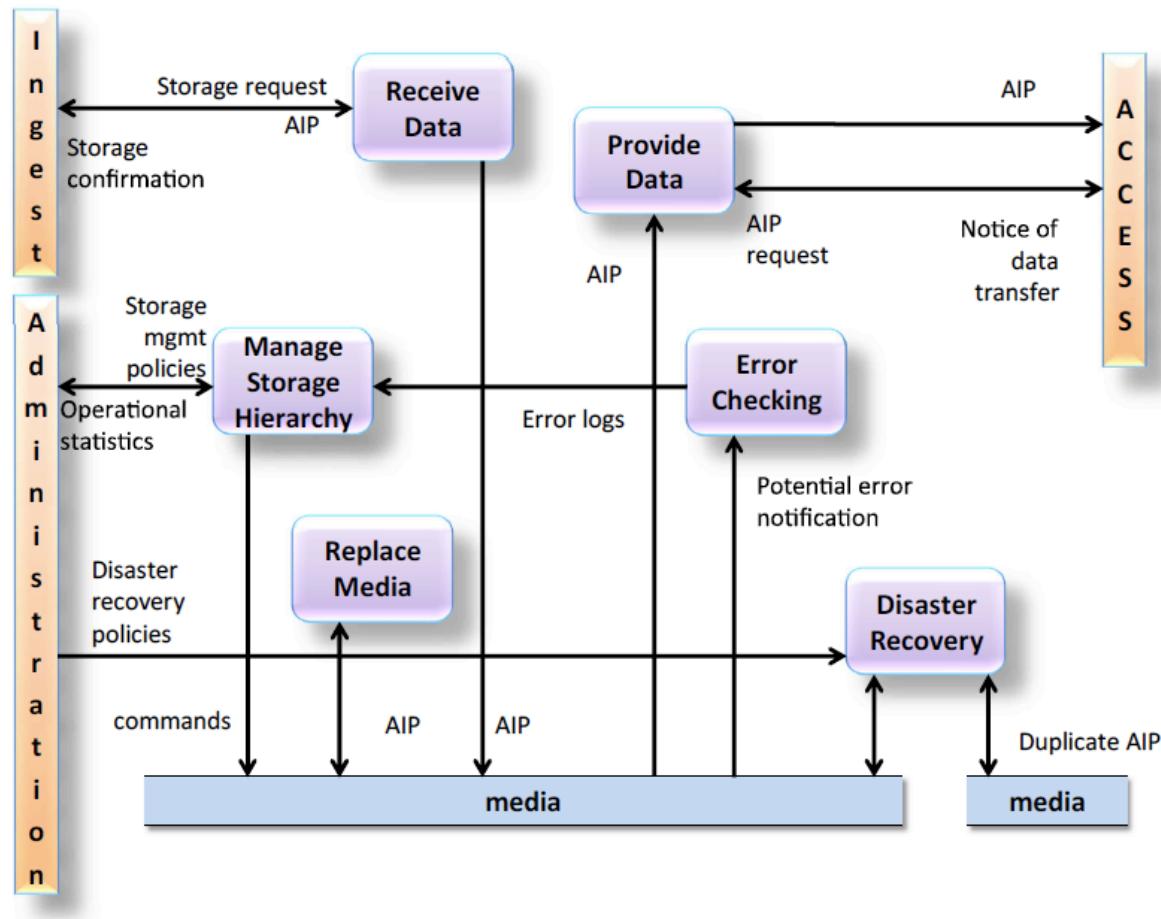
Gli Esperimenti ripetevano, migliorando, misure  
precedenti (ripetere LEP Tevatron & LHC è dura...)

# OAIS

Consultative Committee for Space Systems  
Recommendation for Space Data System Standards  
REFERENCE MODEL FOR AN  
Open Archival Information System  
(Norma ISO 14721:2003)



# OAIS Archival Storage



# Progetti FP7 Esistenti

## SCIDIP-ES

(SCience Data Infrastructure for Preservation - Earth Science)

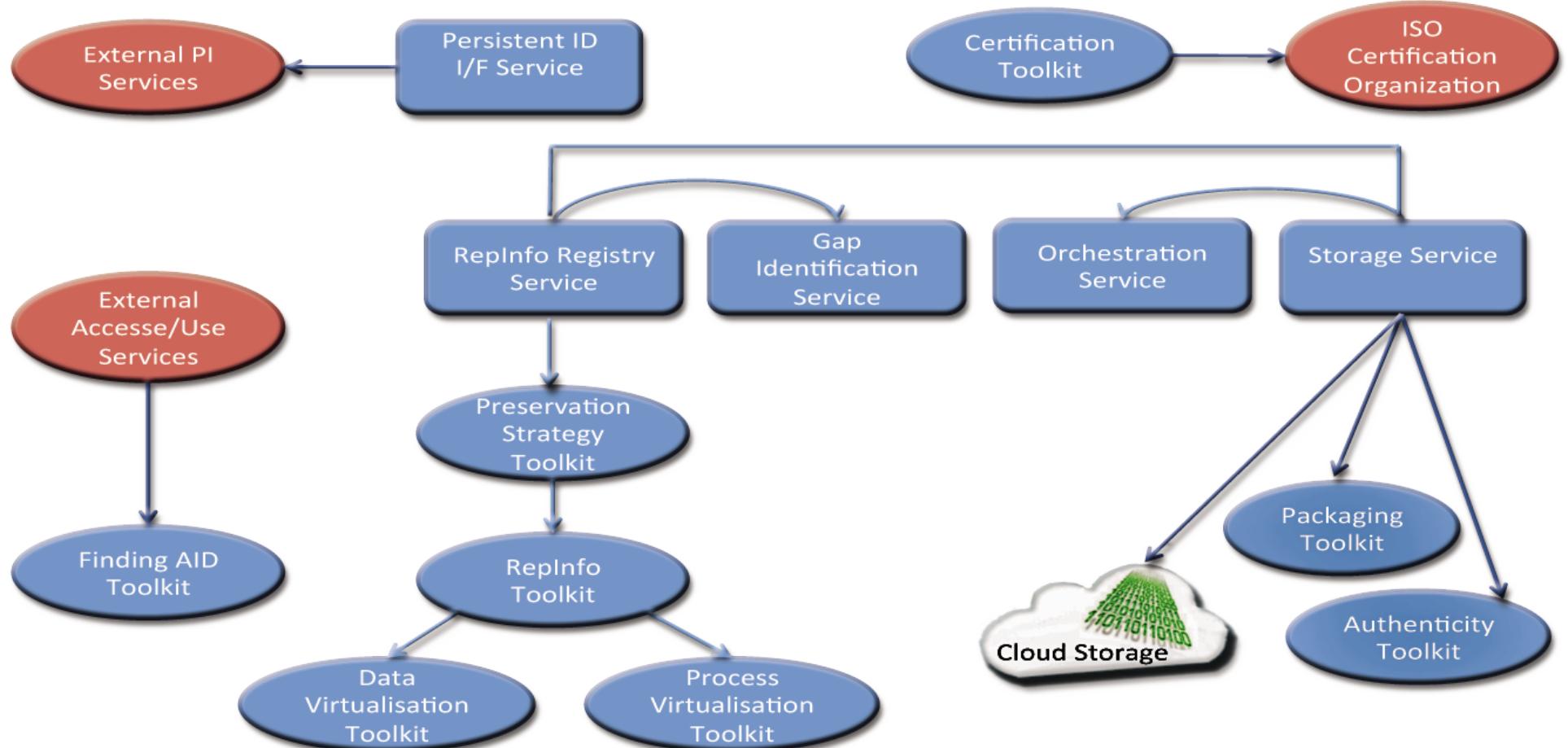
- Progetto principale di Long Term Data Preservation. Eu call INFRA-2011-1.2.2
- It address the issue of building the key information (knowledge) to allow access and understanding of experimental data in a technology independent way such that the preservation is really long term.
- Il progetto vuole realizzare le prime componenti basate su OAIS

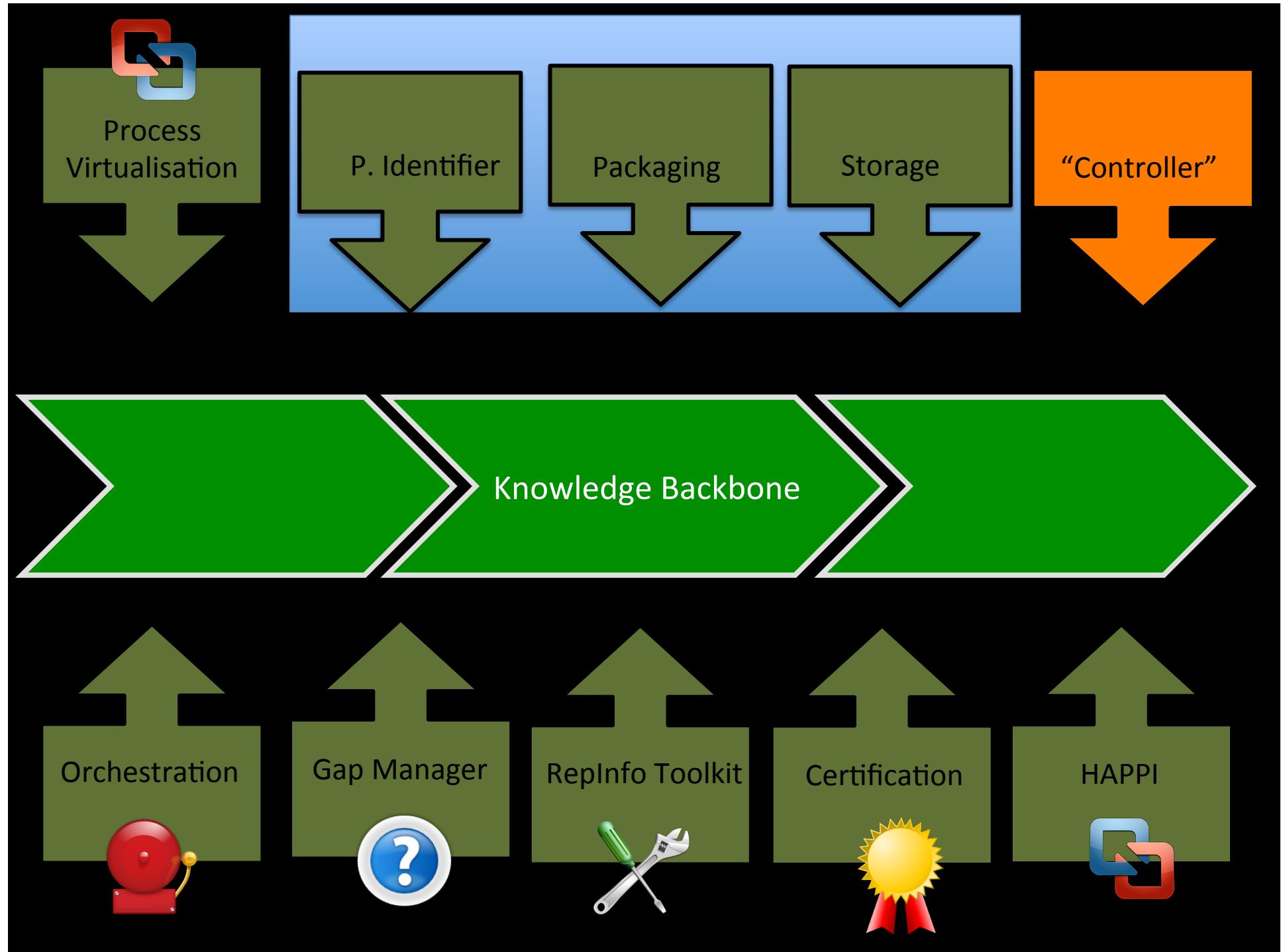
## EUDAT

(EUropean DATa infrastructure)

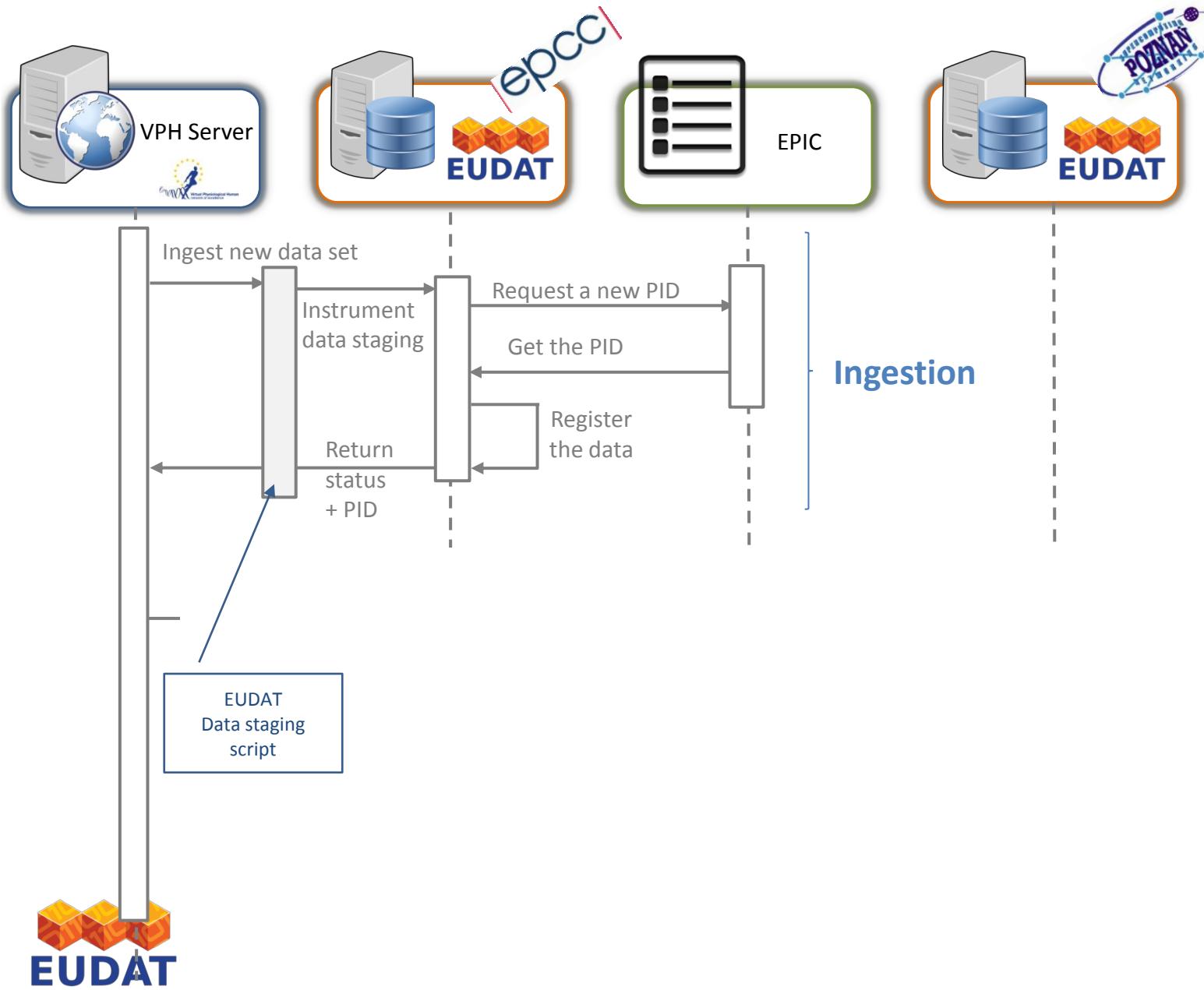
- Progetto per la costruzione di una e-Infrastructure dove i dati siano condivisibili attraverso servizi definiti e procedure standard
- Vuole considerare anche data preservation, ma considera solo bit preservation e poco più

# SCIDIP-ES Services

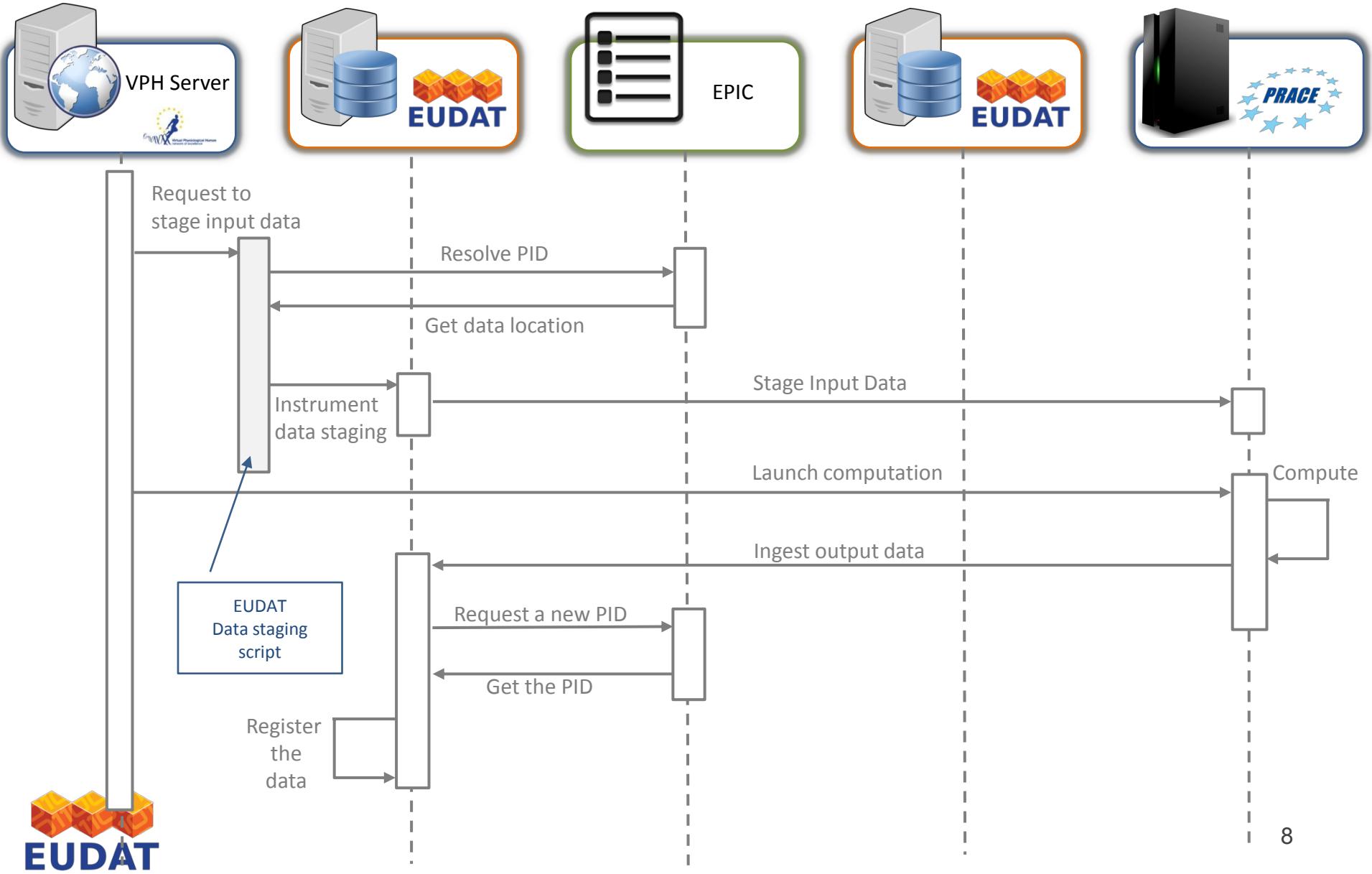




# Ingestion



# Staging



# Sinergie

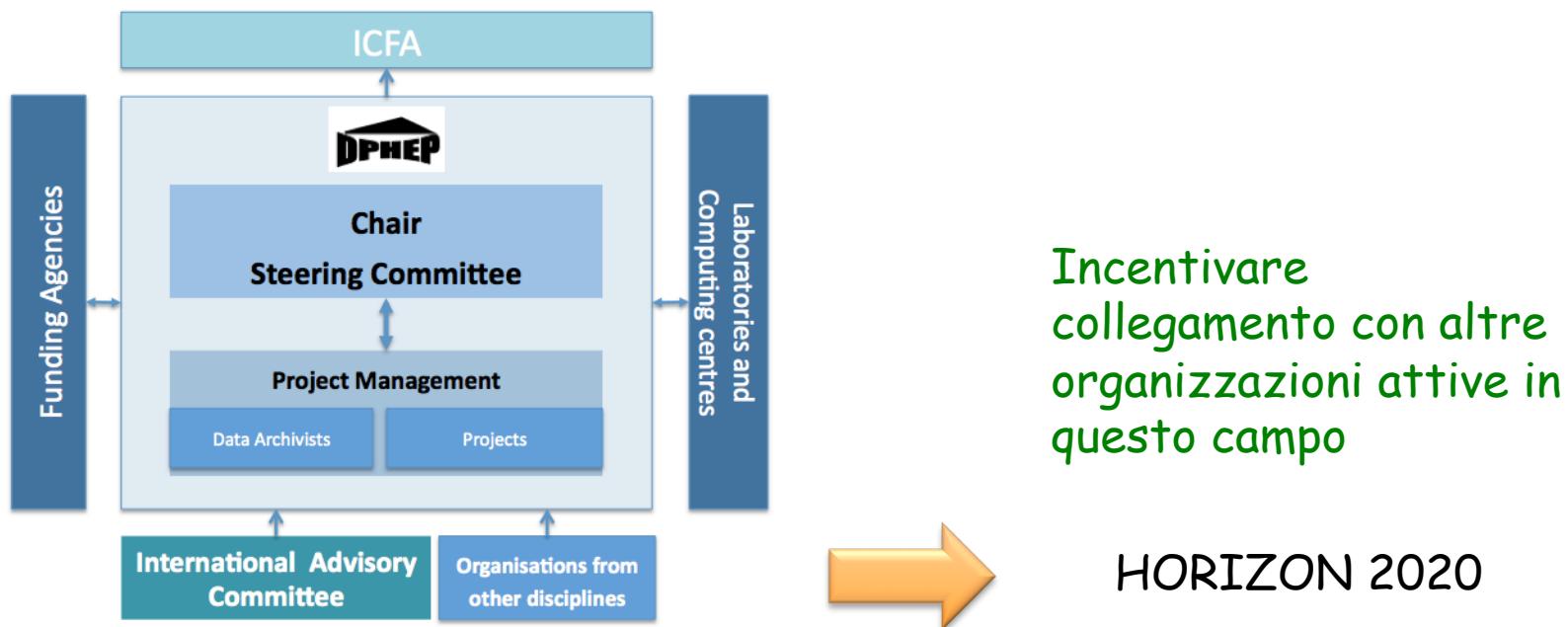
- SCIDIP-ES ha rilasciato le prime release per LTDP basato su OAIS a Novembre
- INVENIO digital library technology
- EUDAT per il data management & bit preservation
- DPHEP per il sistema LTDP per HEP
- APA per lobbying

# DPHEP

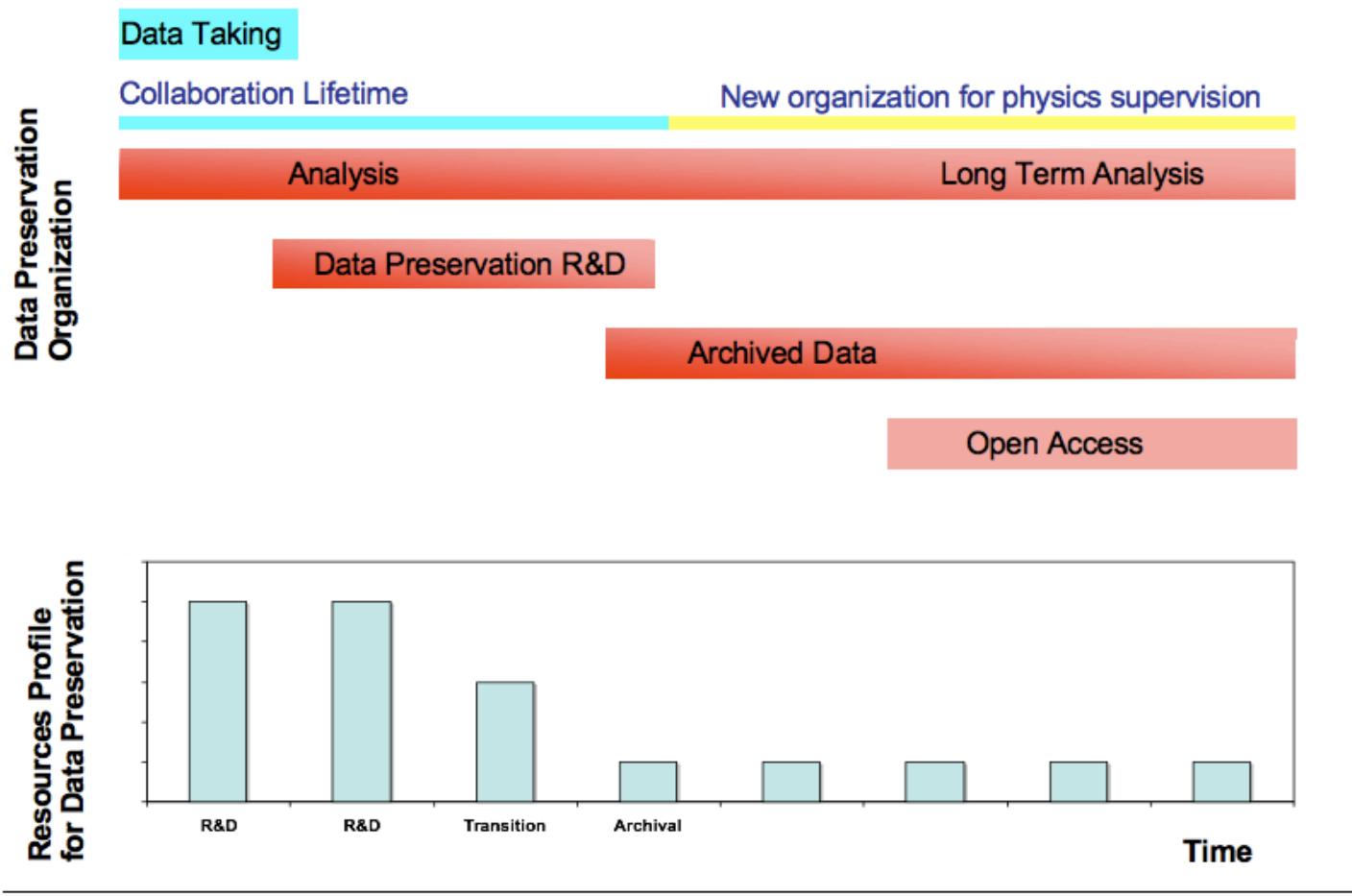
- Pannello dell'ICFA dal 2009

[https://dl.dropbox.com/u/48384809/dataprese/DPHEP\\_2012-05.pdf](https://dl.dropbox.com/u/48384809/dataprese/DPHEP_2012-05.pdf)

- Partenza di una collaborazione internazionale



# Experiment Timeline



# DPHEP Models

Preservation Model	Use Case
1. Provide additional documentation	Publication-related information search
2. Preserve the data in a simplified format	Outreach, simple training analysis
3. Preserve the analysis level software and data format	Full scientific analysis based on existing reconstruction
4. Preserve the reconstruction and simulation software and basic level data	Full potential of the experimental data

<http://www.nature.com/news/lhc-plans-for-open-data-future-1.14244>

Articolo su DPHEP in NATURE il 27/11/2013  
**LHC plans for open data future**

# Nell'INFN

- ALEPH
- AMS
- ARGO
- AUGER-YGS
- CDF
- EAS-TOP
- KLOE
- PAMELA
- MACRO
- STS a LNGS

Esperimenti con persone INFN  
che lavorano o hanno dato  
disponibilità

- Preservazione dei dati
- Preservazione su dell'esperimento
- Data Access System
- Condition Data e Metadata

LHC incomincia ora

# Status

- Eudat: sarebbe utile aprire un canale di comunicazione e proporre degli use per piccoli esperimenti o per il singolo scienziato (Simple Store con Invenio come back-end)
- Scidip: contatto stabilito, middleware disponibile, c'è un problema di semantica da risolvere...
- DPHEP: siamo dentro, ma l'INFN non ha firmato il collaboration agreement
- APA: sarebbe utile iscriversi (come il CERN) anche perchè il direttore governa il tema in Europa (RDA)

# Vision

