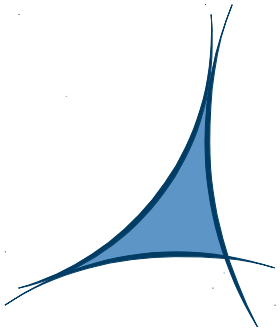

PIC
port d'informació
científica

Data Backbone (ZEPHYR) H2020 proposal arising from the EU-T0 initiative

Manuel Delfino

Presentation to the WLCG MB
15/7/2014

Institut de Física
d'Altes Energies

Centro de Investigaciones
Energéticas, Medioambientales
y Tecnológicas

PIC
port d'informació
científica

Disclaimer

- ZEPHYR proposal is under heavy development, so I can't be too specific

- EU-T0 Collaboration Agreement is being drafted and will be signed in the Fall.

PIC
port d'informació
científica

EU-T0 initiative

- Collaboration agreement to enhance collaboration between data centers
- IN2P3, INFN, STFC, DESY, KIT, CIEMAT, IFAE and CERN
- Not a "project", but a long-term strategy
- All these centers have T0 responsibilities for scientific projects. Astrophysics growing.
- Work together to ease access, reduce costs
- Naturally, this has spawned some H2020 ideas

PIC
port d'informació
científica

Data Backbone (ZEPHYR)

- Proposal targetted at a specific H2020 topic:
  - (7) Proof of concept and prototypes of data infrastructure-enabling software (e.g. for databases and data mining) for extremely large or highly heterogeneous data sets scaling to zetabytes and trillion of objects. Clean slate approaches to data management targeting 2020+ 'data factory' requirements of research communities and large scale facilities (e.g. ESFRI projects) are encouraged.

PIC
port d'informació
científica

Imagining the future

- If you want to guess how different the world will be in 2024, just compare today to 2004.
- "Data Backbone"
  - Hide knowledge of which center hold which data
  - Support automatization of Data Preservation
  - Increase automation
- An example of WLCG inspiration:
  - Xrootd federations

PIC
port d'informació
científica

ZEPHYR approach

- Reports on requirements for 2020+ projects exist (HL-LHC, SKA, CTA, Euclid, etc.)
- Organize work using a loose "architecture"
  - API (what do I call to access my data)
  - Control Plane or Storage Virtualization (mediates between APIs and physical data centers)
  - Local Data Center (how resources plug in)
- Transversal items
  - Data Preservation
  - New network features
  - Security and confidentiality

PIC
port d'informació
científica

ZEPHYR approach

- Don't reinvent the wheel
  - Survey current and emerging technologies
  - Choose some of them and evaluate them
  - Combine some of them for proofs-of-concept and prototype tests
- Loop-back with projects to gain guidance
- End results: Narrow the scope of choices to study as we get closer to 2020+