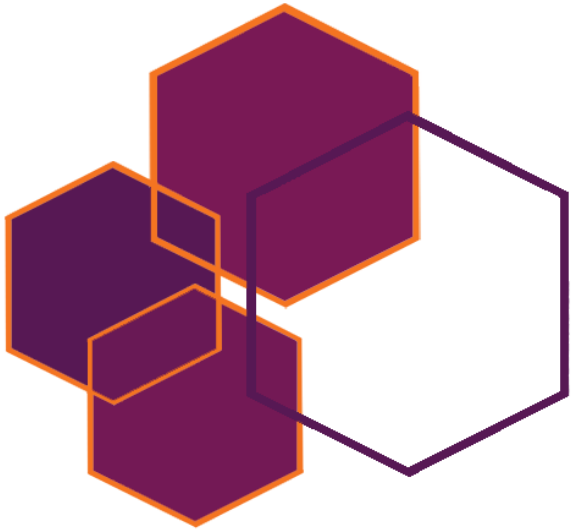


EU-T0

Giovanni Lamanna

LAPP - Laboratoire d'Annecy-le-Vieux de Physique des Particules,
Université de Savoie, CNRS/IN2P3, Annecy-le-Vieux, France

WLCG-OB, 9 May 2014

cnrs IN2P3
The two infinities

EU-T0, Data Research and Innovation Hub

- **EU-T0** is a hub of knowledge and expertise that optimises the investment of the funding agencies in proven e-infrastructure by broadening, simplifying, and harmonising access, driven by well-defined user requirements.

- **EU-T0** aims to build a federated virtual European **Tier 0** data-management and computing centre, implementing the connection and coordination among the major national e-infrastructures.

- **EU-T0** contributes to the implementation of the vision towards a "general purpose European computing and data e-infrastructure for research"

[more agencies are joining]

Visions

- The e-Infrastructure Reflection Group (e-IRG) 2013 White Paper addresses the need of new paradigm for e-Infrastructure for research, able to account for **users needs** and clearly define **roles for all stakeholders**.

- CERN and the EIROforum members have published a vision for the evolution of the European e-Infrastructure aiming to create a **sustainable** IT environment **open to all science communities**. The vision capitalizes on the investments in computing infrastructure made over the last decade, and the facilities in place to support the major European Research Infrastructures (RIs).

- The next generation of services for the **distributed computing and storage infrastructures** have to address the current limitations and profit at best of the important advancement in **Cloud Computing** and in the **CPU architectures**. The next goal is to make available to Europe the next state of the art **distributed infrastructure for Big Data sharing and analysis**.

Rationale

- Considerable overlap between the research funding agencies in Particle, Nuclear, Astroparticle Physics, Cosmology and Astrophysics.

- APPEC promotes opportunities to cooperate in designing new software methods, computing and data processing infrastructures for the current and future major research projects.

- Promoting a large cluster of major research projects and key representatives around e-Science and data management issues.

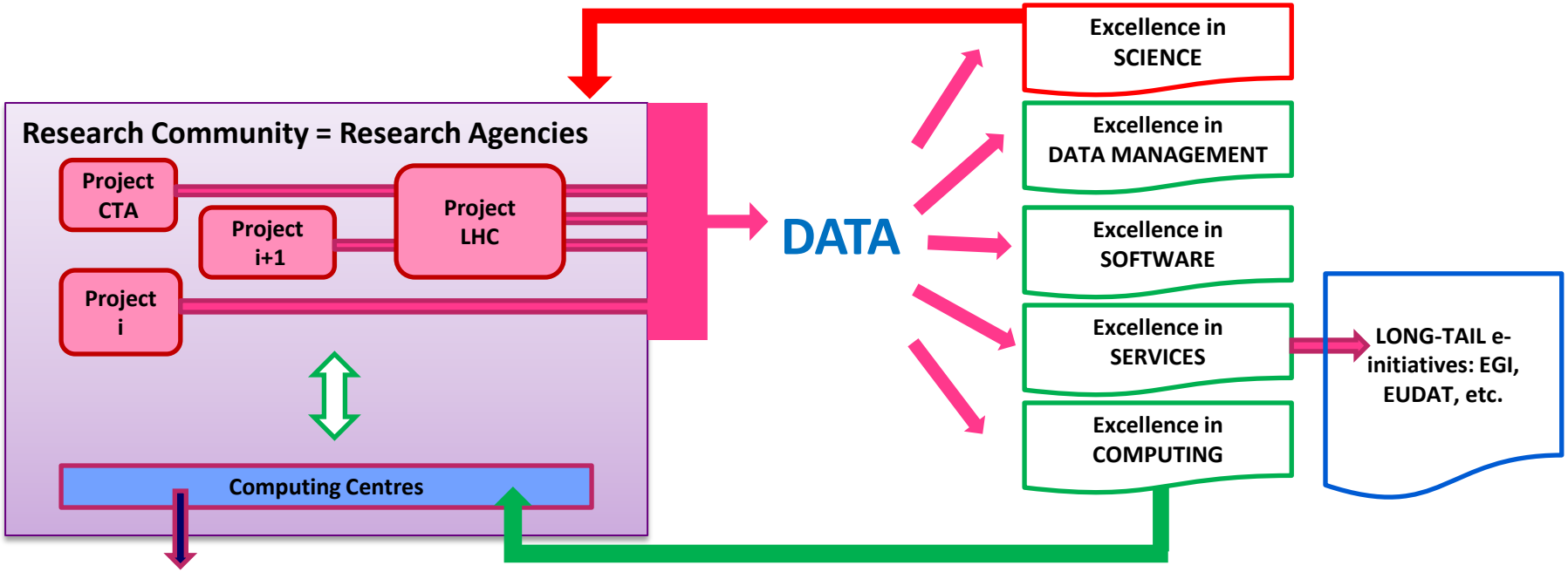
- The EU-T0 federation was launched officially on the 11th February 2014 by some major research institutes in the above mentioned domains: CERN, CIEMAT-ES, DESY-GE, IFAE-ES, IN2P3-FR, INFN-IT, KIT-GE and STFC-UK and the participation is going to be extended to more institutes.

- The approach aims to be a “core” project for a larger multi-domain federation.

Rationale

- Bringing the research communities closer each other to support their needs and: avoid fragmentation and repetitions; increase cross-fertilisation; share standards, expertise and developments; provide and share services; promote outstanding CCs in Europe.

- According to a "data and researchers centric" approach, the EU-T0 e-infrastructure accounts for the "user needs" and the expectations of research communities committed in major ESFRI RIs and ERF.



Rationale

EU-T0

Vision is service-based – set of services to scientists and research projects, built on common underlying infrastructures:

- Advanced, high speed, high capacity international networks among CCs archiving and processing data.
- A functional Authorisation, Authentication and Accounting (AAA) infrastructure.
- Coherent set of data services and tools, capitalizing on the experience of existing major projects

Rationale

EU-T0 IN THE ECOSYSTEM

There is an important need for a number of coordination activities which embraces globally all e-infras. :

- Organization and repositories for sharing services, software, data across communities;
- Recognize shared/common needs for developments and services;
- Efforts to develop new services/tools where needs are identified;
- Policy development (shown to be essential for successful infrastructures);
- Security coordination – policy development and incident handling (CSIRT);
- Collaboration beyond Europe and with more scientific communities;
- Integration with other forms of e-infrastructures including Capability HPC (PRACE) and volunteer computing;
- Coordination/negotiation with industry;
- Creation of a training network to create key skills for the future including “data scientists”

Position statement

A position statement document has been approved by Signatories during a dedicated inter-agency meeting held at CERN on the 11th February 2014.

The Signatories mandate the authors to bring about the following steps:

- To expand the collaboration.
- To agree on the official roadmap for establishing the “EU-T0”.
- To define a detailed work program.

European agencies position statement: towards the “EU-T0” federation.

Authors	Institutes	Approval (Agencies directors)	Institutes
Giovanni Lamanna	IN2P3-FR	Jacques Martino	IN2P3-FR
Donatella Lucchesi	INFN-IT	Ursula Bassler	IN2P3-FR
		Gabriel Chardin	IN2P3-FR
		Fernando Ferroni	INFN-IT
		Antonio Zoccoli	INFN-IT
		John Womersley	STFC-UK
		Antony Medland	STFC-UK
		Joachim Mnich	DESY-DE
		Doris Wedlich	KIT-DE
		Matteo Cavalli-Sforza	IFAE-ES
		Marcos Cerrada	CIEMAT-ES
		Frédéric Hemmer	CERN
Ref.: IA_PS_131209			
Version: 5.0			

Implementation

Currents steps:

- Delegation met EC representatives in Brussels on April 24.

- Vision to be developed over the coming few months:
  - Converging on an explicative document in a few weeks.
  - Taking part to a series of discussions on the e-infra. future in Europe.

- A workshop (28-30 April) discussing some possible immediate H2020 engagements.
  - Discussions are now taking place with EUDAT and EGI.

- Long term goal to create sustainable environment for IT needs of the ERA

- But also potential synergies with US and Asia would be interesting.

Implementation

E-INFRASTRUCTURES

A few pillar projects just started:

- ✓ The EU-T0 data backbone: heterogeneous storage managed in a federated way; interoperable (EUDAT-compatible); fulfilling the real-time ingestion and the archive access requirements;

- ✓ The EU-T0 cloud: extremely hybrid distributed computing architecture, scheduling, virtualizing and configurable for all users; serving also long-tail science.

- ✓ The EU-T0 VRE and software: organization, repositories and provision of services, preserving data, and new software programming test-bed provision across communities;

- ✓ The EU-T0 training: “data scientist” building profile and promoting careers.

- ✓ The EU-T0 pilot: new business model and interaction with private sector for co-developments around big-data services and cloud (Helixnebula).

EU-T0: e-infrastructures background

- The EU-T0 partners are the research agencies owning large computing centers (CCIN2P3, CNAF, GridPP, PIC, DESY-Tier2, KIT-Tier1, … + many Tier2 national CCs) part of the Worldwide LHC Computing Grid (WLCG) project, having successfully implemented a distributed computing infrastructure but also …

- …Supporting large Research Infrastructures (RIs) (some in the ESFRI roadmap) in Astroparticle Physics and Cosmology, such as AMS, AUGER, H.E.S.S., MAGIC, CTA, FERMI, KM3Net, SKA, VIRGO/EGO and future gravitational waves projects, PLANCK, EUCLID, LSST, and in photon science XFEL, etc..

- The EU-T0 hub is built up on gathered resources currently of the order of hundreds of thousands of processing cores, targeting the half a million cores of computing resources in the next few years. EU-T0 archives big, heterogeneous and complex data through storage resources of the order of some hundreds Petabytes, which will grow up at the Exabyte scale already in the next years.

Conclusions

EU-T0 is a new path towards a hub of knowledge and expertise that optimises the investment of the funding agencies in proven e-infrastructures and driven by well-defined user requirements (including those concerning Big Data and Open Access).

Resources are a combination of hardware, software and skills.

Data management and data challenges are part of our ESFRI RIs and ERFs.

The excellence in science is based on excellent data produced by excellent facilities and managed by excellent e-infrastructures. They deserve support in Europe for frontier developments of potential transversal application.

++

Questions?

1) How does (or does not) this benefit WLCG, and what the relationship to WLCG should be?

-> The WLCG project provides global computing resources to store, distribute and analyse the data annually generated by the LHC experiments. Its evolution respond to the LHC projects and requirements.

-> The EU-T0 project aims to promote the proven expertise acquired by the WLCG centers: in which way ?
These centers are involved in other large international projects. The cross fertilizations between projects will help to promote the WLCG services but also to participate to new developments which could be revealed interesting for LHC and viceversa.
EU-T0 would represent the federated platform for development and rationalization of services.
The approach being “data and researchers” centric, will be steered by researchers.

[…]

Questions?

One more consideration:
Some major world-wide future research infrastructures involving even some of all EU-T0 partners' CCs would benefit of a common "desk" where finding inputs from the beginning to adopt the best computing model.
The "data-management" part of any project in Europe is not often well considered (and funded) at the starting phase. The federation would help to focus investments for common developments at the services of all RIs but also for multi-disciplinary usage when researchers are looking for a scientific analysis and data management system for their needs.

[…]

Questions?

2) How will EGI fit into this landscape? Why EU-T0 is different than EGI?
->This would deserve a long discussion and the future is in the hands of the shareholders (funding agencies).

Here I summarize only few common feelings.

a) EU-T0 stays underneath any “transversal initiatives in Europe” which would have as main target the “long-tail research” and/or the provision of SE and CE freely.
b) NGIs support EGI for these targets.
c) EU-T0 partners can today step-down (see Germany) from EGI because FAs have as main objective the support to their RIs. The operations have to be guaranteed independently on any transversal initiative.
d) The Grid usage and the Grid resources in Particle and Astroparticle RIs come thanks to the expertise in the WLCG centers and paid by our FAs. EGI has not effect/role on it.
e) Cloud developments can be envisaged in large cooperation framework also together with EGI and other partners in EIROforum, but… if Cloud is a synonymous of a new technological paradigm for heterogeneous “distributed data and computing infrastructures” system, this should be the core of EU-T0 dev.

[…]

3) How to cooperate with EUDAT?

-> the global data infrastructure purpose cannot be complete without considering also our domain. EU-T0 would have double roles:

a) continuing providing services, expertise and know-how to be extended generally (as EUDAT does already);
b) playing the same role as EUDAT but in our domains where the “data archive ingestion” and “data archive usage, sharing and curation” is a “Big data” issue (almost unique) and when the commitment of multi-decades data preservation and re-use is part of our RI projects.