
Collaboration européenne

Giovanni Lamanna

Chargé de mission informatique – IN2P3/CNRS

Conseil Scientifique IN2P3, 23-24 Octobre 2014

- Missions:
*... stratégie de l'institut; ... promotion international du CC-IN2P3;
...relations avec des partenaires extérieurs...*
- Un point de vue privilégié, les Astroparticules :
CTA data management, Aspera/APPEC and computing
-> demande croissante en informatique; une large communauté scientifique.
- LHC et CERN:
évolution des *computing model* et des défis technologiques toujours de frontière.
- L'interdisciplinarité à l'IN2P3/CC-IN2P3:
Capacités et savoir faire pour servir aussi d'autres communautés.
- Des nouvelles orientations en Europe:
(i) Des défis: *big data, data scientists, data services et computing architectures*;
(ii) Des perspectives pour l'ERA (élaborées par e-IRG, ESFRI, EIROFORUM, ...)

Au-delà de WLCG,

- Certaines agences européennes de financement de la recherche (AF) (e.g. CNRS et CEA) ont également des responsabilités dans d'autres domaines. (*physique nucléaire, des installations de source de lumière, en physique des astroparticules, cosmologie et astronomie.*)
-> profiter de l'expérience déjà acquise dans WLCG et HEP.

Des priorités Européennes et des défis en calcul:

- Ces AF soutiennent des grandes infrastructures de recherche (RI) futures (e.g. ESFRI) dans Astro/Cosmo : **CTA, SKA, KM3Net, EGO** et futurs projets GW, **EUCLID, LSST**; dans le projet de laser d'électrons libres **XFEL**; et d'autres ... (e.g. EPOS, ESS, ESICAT-3D, DARIAH,..)



- Certains de ces projets sont exigeants en termes d'accès et de gestion des données.
- Ils produiront des données du même ordre de grandeur que LHC.
- La plupart sont Observatoires ou Installations accessibles à des très grandes communautés.
- Leur données seront en libre accès.

Des points en commun en Europe

- Rechercher un model durable
 - Ces AF doivent planifier la fourniture en calcul pour les grands projets de recherche.
 - Ces AF doivent faire face à des problèmes de financement.
- Promouvoir le savoir faire de façon formelle
 - Ces AF souhaitent promouvoir et étendre leur infrastructures, en s'appuyant sur WLCG.
 - WLCG est un projet et un nouveau contexte formel de coopération est nécessaire.
- Une évolution naturelle (un rôle obligé?)
 - Les Tier 1 (e.g. CC-IN2P3) ont des ressources qui vont croître à un demi-million de CPU.
 - Ils archivent de grands quantités des données, hétérogènes et complexes (des centaines de Pétaoctets, qui vont grandir à l'échelle des Exaoctets)

Subvenir aux besoins des communautés scientifiques et

- (i) éviter la fragmentation et les répétitions;
- (ii) augmenter la fertilisation croisée;
- (iii) définir des normes;
- (iv) partager l'expertise et les développements;
- (v) fournir et partager des services;
- (vi) promouvoir les grands CC existants en Europe.



Le Consortium EU-T0 a été officiellement lancé le 11 Février 2014 au CERN entre quelques grands instituts et organisation de recherche : CERN, CIEMAT, DESY, IFAE, IN2P3, INFN, KIT et STFC.

Vers la fédération des Tiers 1.

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			



Les signataires chargent les coordinateurs de :

- Développer la collaboration.
- Etablir la feuille de route.
- Définir un programme de travail détaillé.



2. Objectives

The objectives of EU-T0 are:

- I. To build upon the existing, successful and well proven, computing and data processing centres that are owned or funded by the EU-T0 Partners, by enhancing cooperation and harmonisation in respect of all aspects of the infrastructure.
- II. To integrate the computing and data processing centres into a European Tier-0 centre by creating shared services which can be more easily and transparently used by the different Scientific Communities.
- III. To extend and simplify access to the EU-T0 data centres and their services, through a common approach to integrating Authentication, Authorization and Accounting frameworks.
- IV. To achieve a more cost effective provision arising from economies of scale, minimisation of duplication, and adoption of common solutions.
- V. To innovate new services and software tools as required for the benefit of the Scientific Communities. In particular to develop the means to store, move, archive, preserve, and provide open access for Zettabyte scale data sets.
- VI. To integrate with other e-Infrastructures and to work towards convergence of HPC and HTC services.
- VII. To engage with commercial infrastructure providers to develop a workable mixed model of provision.
- VIII. To broaden existing collaborations beyond Europe.



La vision Européenne



The eight EIROforum research organizations combining the resources, facilities and expertise to support European science in reaching its full potential.

Vision: e-Infrastructure for the 21st Century towards a small number of facilities to provide cloud and data services of general and widespread usage; software services and tools to provide value-added abilities to the research communities, in a managed repository: Research Accelerator Hubs.

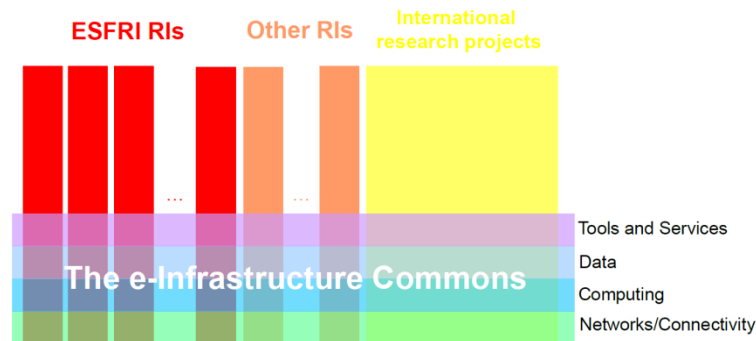
ESFRI

The European Strategy Forum on Research Infrastructures, is a strategic instrument to develop the scientific integration of Europe and to strengthen its international outreach.



To provide strategic advice and guidance on the development of a European e-infrastructure for science and research.

Vision: an open and innovating e-Infrastructure for flexible cooperation and optimal use of all electronically available resources. 2013 White Paper establishes a single e-infrastructure Commons: building blocks exist, coordination is now needed.



E-infrastructures: le scénario Européen

Networks:



The pan-European research and education network that interconnects Europe's National Research and Education Networks (NRENs)

Computing (HPC):



Partnership for Advanced Computing in Europe - a pan-European supercomputing infrastructure, providing access to computing and data management resources and services for large-scale scientific and engineering applications at the highest performance level. An RI in the ESFRI roadmap.

Computing (HTC): missing -> EU-T0 is candidate

[...]



Computing: .. Incomplete, GRID efficient for long-tail science



EGI.eu is a not-for-profit foundation to coordinate and manage the European Grid Infrastructure (EGI) on behalf of its participants.

E-infrastructures: le scénario Européen

Data : .. Incomplete, it is just a project (efficient for long-tail science)



An FP7 project aiming to address the specific challenges of data management, and to ensure a coherent approach to research data access and preservation.

Data : .. Not an infrastructure but an efficient scientific cooperative approach



The Research Data Alliance builds the social and technical bridges that enable open sharing of data.

Vision: researchers and innovators openly sharing data across technologies, disciplines, and countries to address the grand challenges of society.

Data: missing -> EU-T0 is candidate

[...]



Tools and Services: missing -> EU-T0 is candidate to participate

[...]



- Présentation au WLCG MB.
- Présentation au ERF Big Data workshop à Brussels (*après invitation d'APPEC*).
- Présentations à la CE DG-Connect à Brussels.
- Discussions programmatiques en cours avec d'autres partenaires (e.g. PRACE, EUDAT).

- Document S.I. livré (et définition des activités préliminaires).
- Définition de l'interim Consortium Board et interim Executive Board.
- Consortium MoU en préparation

- Extension à d'autres partenaires et projets:
 - i) Les agences T1 NL et Nordic nous répondront; d'autres en discussions.
 - ii) Cohérence bien établie avec APPEC, WLCG.
 - iii) Coordination avec ESFRI en préparation.
 - iv) Coordination avec RDA à l'étude.
 - v) Dialogue avec les Communautés Scientifiques en cours.

- Participation et promotion de projets en réponse aux appels d'offres H2020.

- L'interim Executive Board d'EU-T0 commence à structurer les activités et définir la feuille de route.

Architectures et méthodes des systèmes de gestion des volumes de données approchant le Zettaoctet.

- *Proof of concept* de nouvelle génération de DB.
- Couche globale de virtualisation du stockage qui est logiquement séparé des sites de stockage individuels.
- Simplification et automatisation du déploiement des sites physiques.
- Nécessité de présenter à l'utilisateur une vue intégrée de leurs métadonnées.
- Les défis associés à la nécessité de préserver des données scientifiques.

Les développements proposés font aussi l'objet d'une demande de financement H2020, envoyé le 2/9/2014 (ZEPHYR).

Les développements proposés font aussi l'objet d'une demande de financement H2020, envoyé le 2/9/2014 (**INDIGO-dataCLOUD**).

Exemple de collaboration large avec diverses communautés, initiatives (EGI), autres partenaires non limité a EU-T0, Secteur privé, Cloud commerciale

- Coordination des travaux pour développer et déployer des services virtualisés.
- Collaboration avec d'autres projets dans ce domaine.
- Projet pilote de l'intégration des ressources de Cloud commerciaux.
- Vers un tissu de calcul sous-jacente simple qui peut être plus facilement adaptée aux différentes communautés scientifiques.

HTC & HPC, diverses démarches complémentaires:

- (i) Au niveau national les coopérations HTC+HPC s'intensifient (en France: projets CPER? nouvelle liste TGIR!? Des actions/programmations en commun avec Michel Daydé sont indispensables).
- (ii) EU-T0 propose un soutien (complémentaire) aux initiatives des coopérations entre communautés scientifiques:
- Projets H2020 : Cluster de projets ESFRI en Astron. & Astrop. (ASTERICS); projets Centres d'excellence.
 - Actions de type « Training » en programmation parallèle et GPUs en cours d'élaboration.
 - Des exemples au CERN : projet GEANT V, *HEP software foundation*.
- (iii) Coopération avec PRACE:
- En février 2015 premier appel d'offre en commun **PRACE + EU-T0** (temps calcul + *data postprocessing archive, access and management*).
 - Planification des rencontres entre les managements des deux Consortia pour la formulation d'une possible roadmap commune en Europe.

En EUROPE:

« Plus en générale il y a un point très fort sur la persistance des infrastructures numériques et l'engagement de la Commission Européenne sur leur processus de développement, compétitivité et durabilité.

PRACE et **EU-T0** se basent sur Centres (HPC et HTC) existantes, dont les agences ont la propriété: entité, infras. et services persistantes; différent d'autres initiatives comme EGI, dont le modèle est basé sur le courtage des ressources d'infrastructure fournis par des tiers (qui?). EUDAT est juste un projet.

EU-T0 et PRACE comptent sur les investissements et l'engagement des états membres qui affirment de vouloir doter l'Europe des moyens pour soutenir l'activité de recherche de nombreux scientifiques, et beaucoup d'ESFRI RI. »

Demandes des nouveau
Observatory and Facilities pour
garantir l'accès et le partage
sécurisé et autorisé aux données,
services et infrastructures.

Service pour la federation des T1.

La CE finance en H2020 un/deux
projets en coopération avec
eduGAIN et quelques partenaires
d'EU-T0.

On vise a participer aux
développements techniques et la
mise en opération de *proof-of-
concept*.

- Coordination du déploiement des systèmes de ***Autorisation Authentication and Accounting*** pour simplifier l'accès pour les communautés scientifiques diverses.
- Politique et protocoles internationaux
- Développement d'outils et services pour compléter ou adapter les infrastructures européennes ou mondiales AAA aux divers projets.
- Le bénéfice d'un «single sign on» pour des ressources multiples.

Deux niveaux:

- Le patrimoine des services pour la gestion et le calcul de Big Data: e.g. INVENIO, CERNVMfs, RUCIO, DIRAC, PANDA, XROOTD, dCACHE, iRODS, OperationPortal, OpenAIRE, etc...
 - Les services des niveau plus haut pour les utilisateurs: e.g. applications pour l'accès aux données, service de repository et développement logiciel, publications d'articles et données/metadonnées associés.
 - La préservation des données: méthodes, couts et services.
- Coordination et collaboration dans le développement, la fourniture et le déploiement d'un portefeuille de services
 - Construction d'un **Virtual Reserach Environment** de recherche personnalisable.
 - Services de gestion de flux de données, d'accès des utilisateurs à des logiciels, des archives et des systèmes d'analyse scientifique.
 - Plate-forme numérique concernant les publications et des référentiels aux données et aux logiciels d'analyse.

Ma vision personnelle sur diverses démarches:

- “Certification de la fédération” des Tiers 1 d’EU-T0 pour l’archivage pérenne et sécurisé des données.
- “Intensifier la vocation multidisciplinaire” du CC-IN2P3 avec des ressources financières plus importantes et d’origines différentes mais durables.
- « Model économique » de prestation des services pour les données: vers le “Tier 0” Européen.
- Synchroniser roadmaps TGIR et ESFRI:
 - Soutien politique de chaque pays membre d’EU-T0 dans ESFRI, e-IRG et CE. Possibilité d’un financement ad hoc H2020 en 2016-2017.
 - La roadmap ESFRI devrait voir sortir PRACE de la branche IT. Réfléchir (avec HPC) pour proposer une nouvelle infrastructure ESFRI – IT ??

IN2P3 (et CEA): valoriser les parcours des chercheurs dans le “data management” et participer a la promotion des profils “*data scientist*” (message pour la Section 01)

IN2P3: promouvoir les synergies entre équipes de recherche et le CC-IN2P3 depuis le premier moment d'un projet (*big data*).

IN2P3, CNRS: création des groupes de travail avec les chercheurs en informatique si “stimulés” par nos “*use cases*”

IN2P3: confirmer le soutien a EU-T0 et sa coordination et promouvoir la structuration de la contribution Française:

- réorganisation des acteurs du calcul (*e.g. synergies avec FG et LCG-FR et plus de coordination en soutien d'EU-T0, déjà envisagés mais redéfinitions des stratégies et des priorités sont nécessaires*).
- cohérence avec la liste TGIR.
- stratégie pour le CC-IN2P3.
- soutien dans les interfaces avec le Ministère (voir avec Mark Asch), avec la CE (voir nos contacts a Brussels, mais aussi DG-Connect), avec e-IRG (voir Ministère), avec ESFRI, avec RDA.

CNRS: soutien en cohérence avec l'évolution du paysage HPC/HTC (voir avec Michel Daydé) mais aussi partager la démarche EU-T0 et la promouvoir auprès des communautés scientifiques.