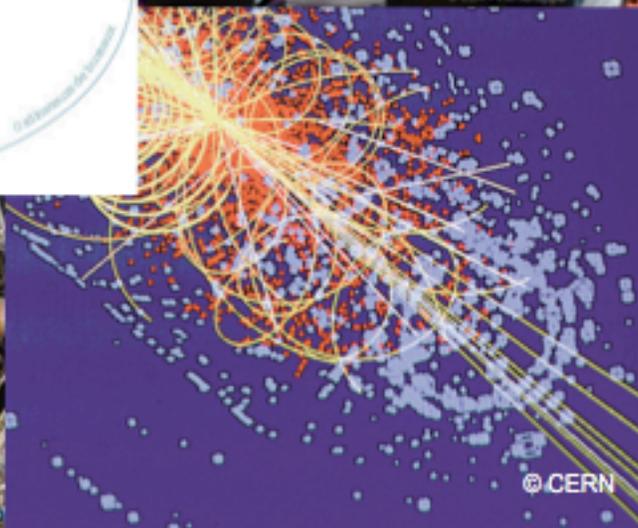
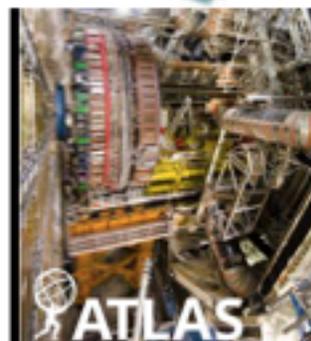
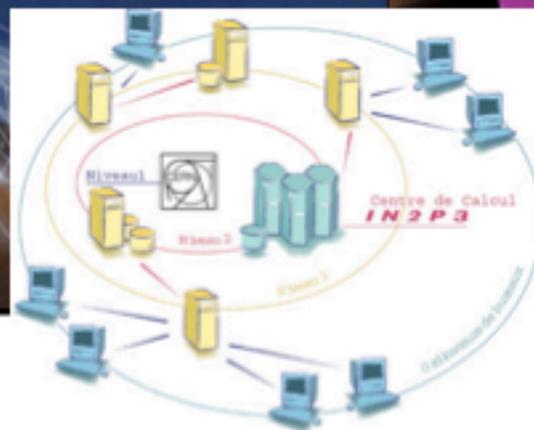
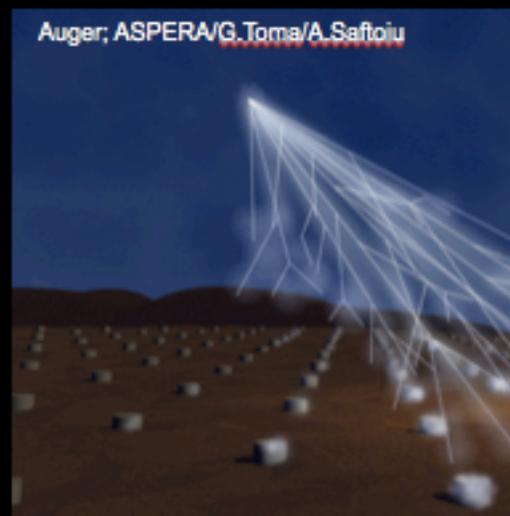


Vision de la direction IN2P3 sur l'informatique et le calcul



Volker Beckmann
CNRS / IN2P3

Journées Informatique l'IN2P3 - IRFU

Défis Informatique



Pourquoi un nouveau domaine calcul & données à l'IN2P3



TECHNOLOGIES CLÉS
Préparer l'industrie du futur

2 0 2 0



DGE

3 Valorisation et intelligence des données massives

LOISIRS & CULTURE
ÉNERGIE, MOBILITÉ, NUMÉRIQUE
ENVIRONNEMENT, HABITAT, SANTÉ ET BIEN-ÊTRE, SÉCURITÉ
ALIMENTATION

> Correspond à une technologie clé 2015

MOTS CLÉS
Données massives, Big data, mégadonnées, analyse, stockage, données personnelles, prédiction, cloud computing, cybersécurité, capteurs, visualisation.



4 Modélisation, simulation et ingénierie numérique

LOISIRS & CULTURE
ÉNERGIE, MOBILITÉ, NUMÉRIQUE
ENVIRONNEMENT, HABITAT, SANTÉ ET BIEN-ÊTRE, SÉCURITÉ
ALIMENTATION

> Correspond à une technologie clé 2015

MOTS CLÉS
simulation, modélisation, prévention, prédiction, circulation, visualisation, intégration, déluge de données, hétérogénéité de données, logiciels, marché concentré



Pourquoi un nouvel DAS* informatique?

- toute science (IN2P3) repose en grande partie sur l'informatique
- nouveaux défis scientifiques (LHC run 3, LSST,, ...) exigent une nouvelle approche dans l'informatique
- Attirer des ingénieurs et chercheurs en informatique
- Informatique : service → recherche
- Défis d'infrastructure
- Nouvel accent IN2P3 sur calcul et données

* DAS : directeur adjoint scientifique

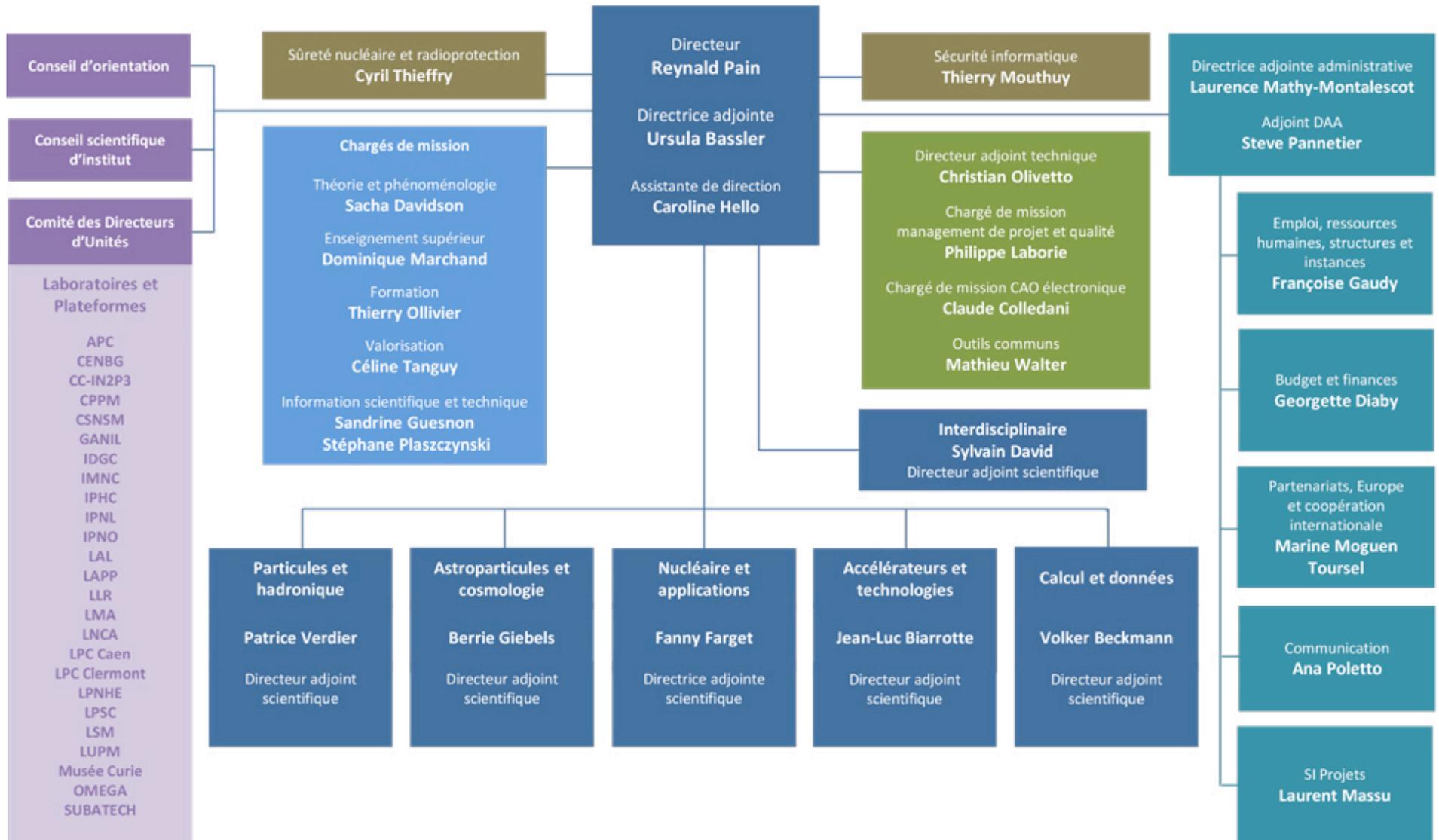
Objectifs principaux



- Recherche dans le domaine informatique appliquée
- Coordination recherche dans l'IN2P3 et avec partenaires (nationaux, internationaux)
- Formation, master, doctorat, HDR
- Infrastructure calcul (Grille / EGI, CC-IN2P3)
- Organisation optimale (meso-)centres de calcul
- Publication: « Computing & Software for Big Science »



Organisation



Organisation scientifique



Directeur
Reynald Pain



Directrice adjointe
Ursula Bassler



Directeur A. Technique
Christian Olivetto



DAS Interdisciplinaire
Sylvain David

Particules et
hadronique

Astroparticules
et cosmologie

Nucléaire et
applications

Accélérateurs et
technologies

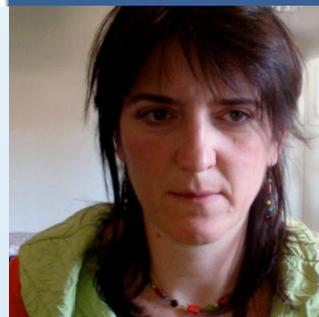
Calcul et
données



Patrice
Verdier



Berrie
Giebels



Fanny
Farget



Jean-Luc
Biarrotte



Volker
Beckmann

Opportunités



- Coordination et animation
- Visibilité recherche informatique
- Attractivité labos IN2P3 domaine informatique

- Formation
- Soutiens financiers
- Postes permanentes / CDD
- Mais : aucune augmentation du budget global IN2P3

Hors périmètre DAS calcul



- Service informatique comme un service technique labo
→ DAT* Christian Olivetto



- Soutien infrastructure calcul labo → SBNA*

* DAT: Directeur Adjoint Technique

* SBNA: Soutien de base non affecté

Objectif

Apporter les meilleures solutions informatiques
pour les projets scientifiques de l'IN2P3



Objectif



Apporter les meilleures solutions informatiques pour les projets scientifiques de l'IN2P3

Cela inclut:

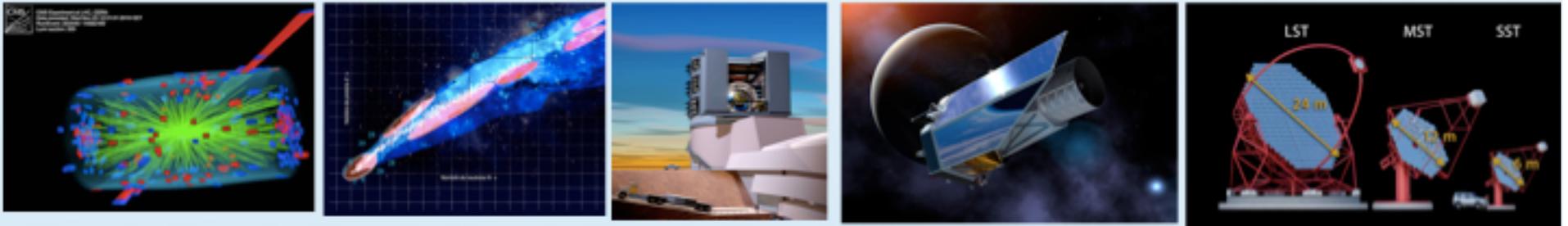
- Viser le Open Source, open access (e.g. prérequis H2020)
- Planifier au niveau IN2P3 – ne pas se limiter au niveau local

Objectif

Apporter les meilleures solutions informatiques pour les projets scientifiques de l'IN2P3

Défis:

- LHC run 3 (2020 – 2023)
- Physique nucléaire, Spiral2, etc.
- LSST (2021 -), Euclid (2021 -), CTA (2018 -), ...



Solutions



Apporter les meilleures solutions informatiques pour les projets scientifiques de l'IN2P3

Besoins:

- RH! Motivation, formation, communication
- Infrastructure: local, régional, national, international
- Organisation: RI3, LCG, France Grilles, direction IN2P3, COCIN¹, EGI², ...



¹ Comité d'Orientation du Calcul Intensif du CNRS

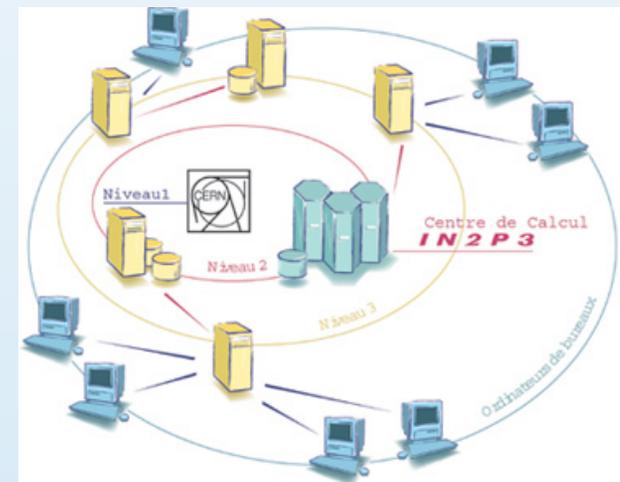
² European Grid Infrastructure

- Formation continue
- Soutien de la recherche en informatique, nouveaux projets
- PhD, HDR → attirer étudiants
- Améliorer la communication entre labos
- Computing Journal → nouvelle motivation pour collaborations nationales et internationales (présentation Michel Jouvin, jeudi matin)



Infrastructure

- Hardware centralisé et expertise locale
- Construire sur la structure locale: ASR, support local
- Suivre les compétences
- Grille / Cloud → cloud fédéré
- Intégration du HPC*



* HPC: calcul à haute performance

Focus!

- Expertise est la clé
- Se focaliser sur les compétences locales (GPU, ML*, Cloud, ...)
- Engagement optimisé des RH (et du budget)
- Partage des compétences
- Collaborations: apprendre à partager les ressources IN2P3

- Pas nécessaire que chaque labo ait toutes les compétences et le hardware, mais au besoin chaque labo doit avoir accès aux compétences et hardware disponibles au sein de l'IN2P3

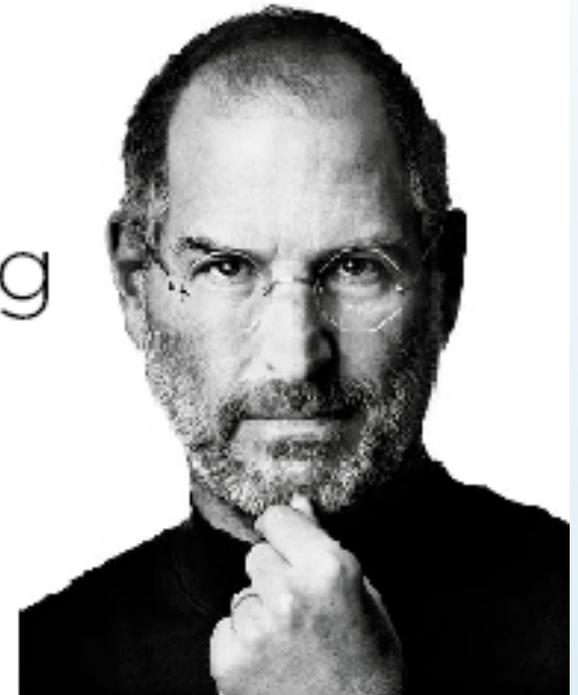
* ML: Machine Learning



Focus!

“Deciding what not to do is as important as deciding what to do.” *

- Steve Jobs



* Décider ce qu'il ne faut pas faire est aussi important que décider ce qu'il faut faire

Organisation



- RI3
- Réunion en janvier: chefs de SI, responsables informatiques de grands projets, LCG, France Grilles, IdGC, ..., IN2P3

Notre place en Europe:

- EGI, EUDAT, WLCG, EU-T0, RDA ...



EGI: European Grid Infrastructure,
WLCG: Worldwide LHC Computing Grid,
RDA: Research Data Alliance

Organisation



France:



- GENCI: grands centres de calcul nationaux
- RENATER: réseau, outils de communication

Besoin d'un équivalent français de EOSC*?

- Combiner France Grilles, LCG, IdGC, ...
- Améliorer accès, partage et maintenance de grille & cloud
- Améliorer la position française (IN2P3) par rapport aux partenaires européens

* EOSC: European Open Science Cloud



Data Management Plan (DMP)



- Les données sont le fruit de nos expériences et le fondement de tout résultat scientifique
- Leur gestion représente un coût
- Conserver à long terme toutes les données, produits, logiciels,... : c'est très (trop) cher
- Perdre des données irremplaçables avant leur analyse : c'est un désastre

- Planifier quoi faire des données, quand le faire et par qui (what/when/who) économisera de l'argent et des ETP
- Les DMP sont requis par les agences de financement: H2020, CNES, IN2P3...



Cycle de vies des données



- **Créer ou collecter** des données (Create)
- **Documenter** les données (Document)
- **Accès** aux données et utilisation (Use)
- **Stocker et sauvegarder** des données (Store)
- **Partager** les données (Share)
- **Conserver** les données (Preserve)

Credits: Jones S., 2015

DMP pour les “projets” utilisateurs du CC-IN2P3



Formulaire

- Info administratives
- Description données
- Métadonnées et documentation
- Responsabilité
- Ethique, cadre légal (droits de propriété)
- Stockage sécurisé et sauvegarde
- Partage des données
- Préservation a long terme (archivage)
- Budget

« Pourquoi est-ce important ? »



Feuille de route

- Nouvel accent IN2P3 sur calcul et données
- Garder l'objectif bien en tête, se focaliser
- Partager expertise et infrastructure
- Think IN2P3, think European!
- Regrouper cloud & grille pour améliorer service, maintenance, impact
- Data Management Plan pour tous les projets
- Réunion janvier 2017
- Comité d'Orientation 2017 → prospective scientifique
- **Nous avons besoin de vous pour définir le future !**



V. Beckmann