



[www.cnrs.fr](http://www.cnrs.fr)

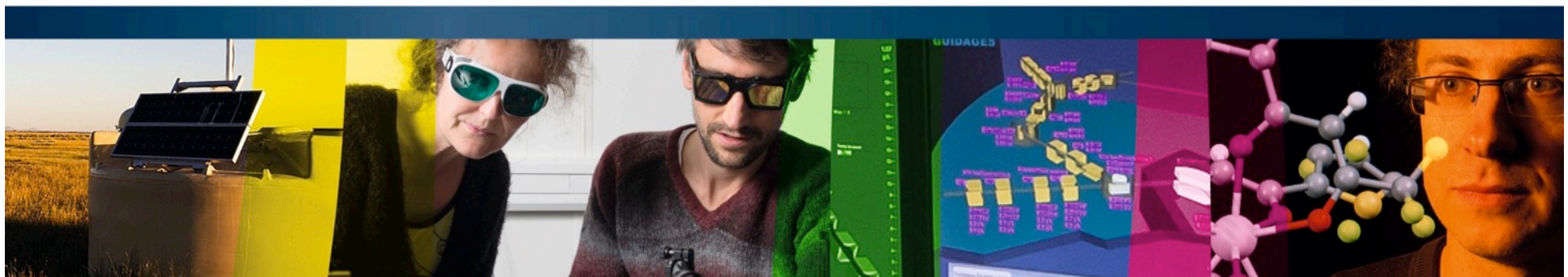
# ANF Mathrice

Mardi 4 décembre 2018

CIRM

Christophe Berthon

Chargé de mission « Données et Calculs »





# Big Data & HPC : des enjeux stratégiques



## **Modélisation et simulation**

### **Exploitation des données (« Big Data »)**

#### **Au cœur des grandes avancées de la recherche scientifique**

➤ Génome humain, boson de Higgs, évolution du climat, risques naturels, pollution atmosphérique, environnement...

#### **Nombreux autres défis scientifiques**

➤ Structure de l'univers, astrophysique, neuroscience, combustion, sismologie, climat, biologie et recherche médicale, matériaux, ....

#### **Enjeu stratégique de compétitivité et d'attractivité internationale**

multiples champs disciplinaires ; importantes retombées socio-économiques



# Big Data : Quelles spécificités?

## Volume

- PetaBytes : un défi pour les architectures de stockage

## Variété

- Diversité des contenus, des formats et des données

## Vitesse

- Un défi pour les réseaux
- Nouveaux modèles pour traiter des flux de données

## Pertinence / Qualités

- Données souvent “write-once, read-many”(WORM)
- Parfois couteuse à acquérir
- Souvent fortement parallélisable

## Valeurs des données





# Besoins immenses autour des infrastructures de données

## Communautés scientifiques avec des besoins / compétences bien établies

- Sciences de l'univers
- Physique des hautes énergies
- Biologie
- ...

## Foisonnement d'initiatives plus ou moins coordonnées

- Multitude d'acteurs
- Multiplicité des sources de financements
- Multiplication des infrastructures
- Coûts induits élevés

## Besoin structuration nationale et locale





# Calcul + données : coûts croissants à maîtriser

## Calcul intensif

- Coordination au moins au sommet de la pyramide du calcul (Tier-0 / Tier-1 / Tier-2, EQUIP@MESO,...)
- Evolution technologique + consommation énergétique croissante + adaptation des codes + compétences + support



## Données

- Explosion des besoins et des demandes non coordonnées (CPER) même si certaines communautés sont structurées (physique des hautes énergies, sciences de l'univers, bio, ...)
- Conforter compétences + support

## Impact sur l'organisation de la recherche



## Mission Calcul – Données (MiCaDo)

### Contexte

- Développement modélisation / simulation
- Avancées scientifiques significatives
- Explosion des besoins de puissance de calcul
- Explosion des volumes de données

### « Discipline » transverse

- Par nature très inter-disciplinaire
- Pas de réelle “appropriation” par chaque communauté
- “Maturité” très différente selon les disciplines





## Mission Calcul – Données (MiCaDo)

- Définition et mise en œuvre d'une politique globale et cohérente du CNRS sur les objets à coût significatif (y compris en ETPT)
- Pilotage ou participation au pilotage des ressources dédiées
  - IDRIS, CC-IN2P3
  - France-Grilles
  - Maison de la Simulation
  - Gricad
  - ...
- Cohérence de la politique du CNRS avec les “objets” nationaux (GENCI, Renater, Infranum,...) ou internationaux (PRACE,...)
- Aider les DSR à construire une stratégie conjointe avec nos partenaires





# Livre Blanc sur le Calcul Intensif au CNRS



- Lancé en Avril 2011 sur la base d'un questionnaire
- Contribution du CNRS à la réflexion lancée par GENCI sur les besoins des communautés et sa stratégie pour 2012-2016 (renouvellement IDRIS en 2012)
- Panorama du calcul intensif dans les Instituts : pratiques, besoins, verrous, cartographie, aspects recherche, ...
- Certains des instituts (e.g. INC, INP) mènent régulièrement ce type de travail
- Travail finalisé fin 2012 avec pour objectif de mener des réactualisations régulières






# Livre Blanc sur les données au CNRS



- Lancé en Septembre 2017 sur la base d'un questionnaire
- Contribution du CNRS à sa stratégie pour 2018-2012
- Panorama de l'utilisation des données dans les Instituts : pratiques, besoins, verrous, cartographie, aspects recherche, ...
- Travail finalisé début 2018 avec pour objectif de positionner le CNRS dans une actualité très mouvante



## Éléments clés de la stratégie du CNRS

- 
- **Ne plus dissocier HPC de l'analyse et valorisation des masses de données**  
issues des simulations numériques (climat, fluides turbulents,...), grands instruments (ITER, plateformes génomiques ...) et grands systèmes d'observation au sol (sismologie et géodésie) et dans l'espace (imagerie et interférométrie)...
  - Calcul + data intensif pas uniquement problème de ressources un **changement de paradigme** dans la recherche scientifique :
    - Plus d'inter/pluridisciplinarité (informatique, maths et autres disciplines)
    - Vision holistique des Infrastructures calcul / données / grands instruments / plateformes expérimentales / systèmes d'observation
  - **Rationaliser le déploiement des infrastructures, coordonner les demandes**

**En s'appuyant stratégie nationale / de site autour de défis scientifiques et maîtrise des coûts**



## COCIN (Comité d'Orientation pour le Calcul Intensif)



- Créé en décembre 2010
- Réflexion collective sur les besoins, la structuration et les évolutions en calcul intensif au CNRS
- Prospective sur les besoins des différentes communautés, proposition de maintenance et de développement coordonné des moyens / ressources liées au calcul intensif, en particulier pour l'IDRIS.
- Dix personnalités scientifiques désignées par chacun des instituts du CNRS plus le Directeur de l'IDRIS.
- Le président et directeur désignés par le Président du CNRS



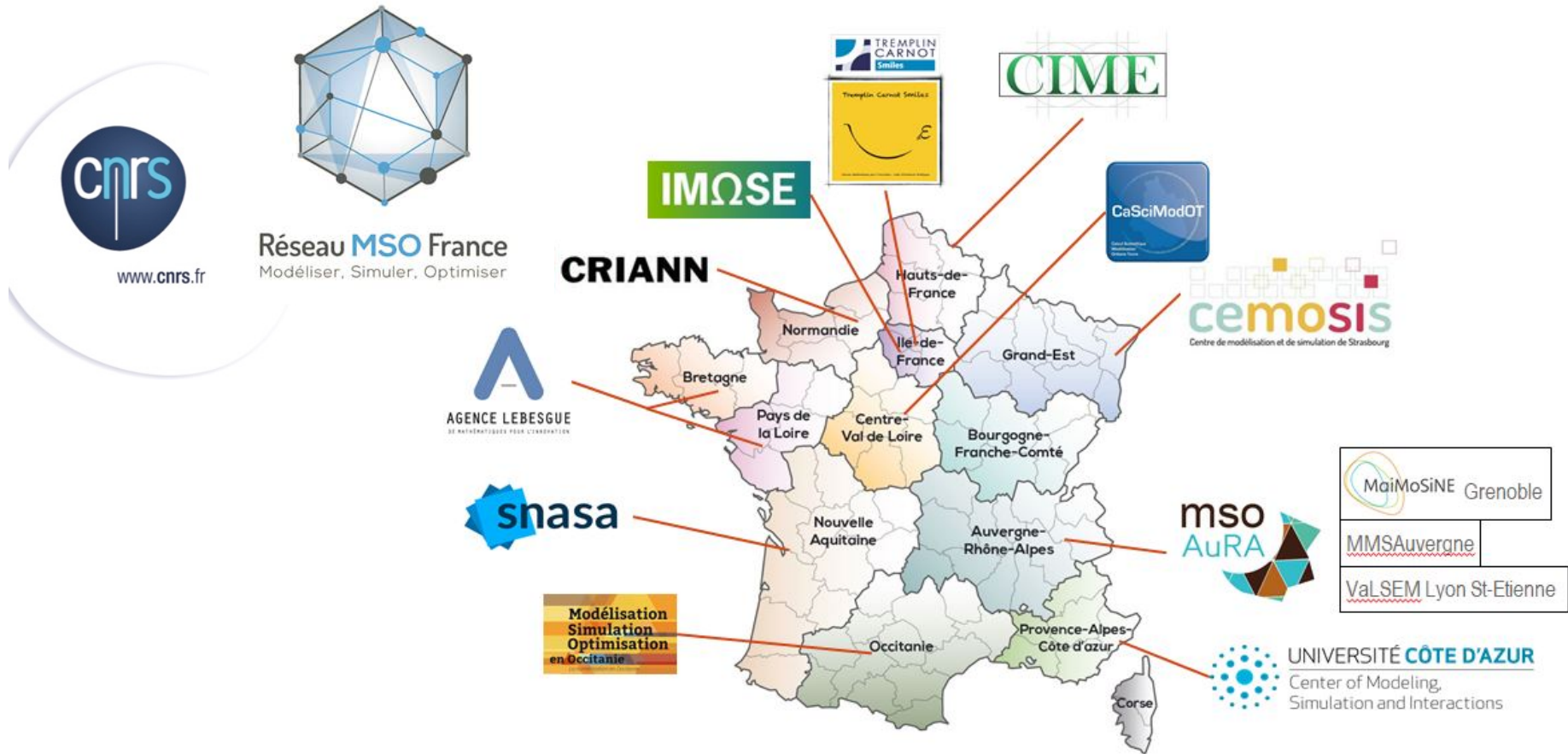
Agence pour les mathématiques  
en interaction avec l'entreprise et  
la société



- LABoratoire d'EXcellence créé en 2011
- Meilleure interaction des mathématiciens avec les entreprises
- Proposer et soutenir des programmes, en formation et recherche
- Visibilité des opportunités qui existent en mathématiques
- membre fondateur du réseau européen EU-Maths-IN (European service network of mathematics for industry and innovation)
  
- Actions : PEPS, SEME, FEM, prix de thèse AMIES,...



# Le réseau MSO : **M**odélisation **S**imulation **O**ptimisation





- Créer en 1994 : période de transition du “calcul vectoriel” vers le “calcul parallèle”
- Information : un bulletin trimestriel Bi-ORAP
- Formation : favoriser les actions de formation de haut niveau (Ecoles d’été, Ecoles de GdR, ...)
- Europe : un relais entre les utilisateurs du HPC et des organismes tels que GENCI en France ou PRACE
- Ouverture : organisation de 2 forum par an