



ISCD: institut des sciences du calcul et des données

Pascal Frey

Sorbonne Universités, Université Pierre et Marie Curie

Éléments de contexte

- le **processus de spécialisation thématique**, consécutif à l'explosion des connaissances et des informations, a modelé les structures de nos établissements: « **verticalisation** » des laboratoires dans des UFR attachées à une discipline, **hiérarchisation** de l'organisation;
- la plupart des disciplines connaissent une évolution profonde et irréversible et deviennent désormais des **sciences « computationnelles »** qui sollicitent de plus en plus le calcul scientifique et l'analyse de données;
- il y a une demande croissante de **service** et de **support** à l'analyse et au traitement de données;
- l'**adaptation** de modèles, de méthodes ou d'algorithmes développés dans un contexte différent devient essentielle à l'activité des scientifiques du domaine et représente une **part substantielle** de leur activité.

Contexte

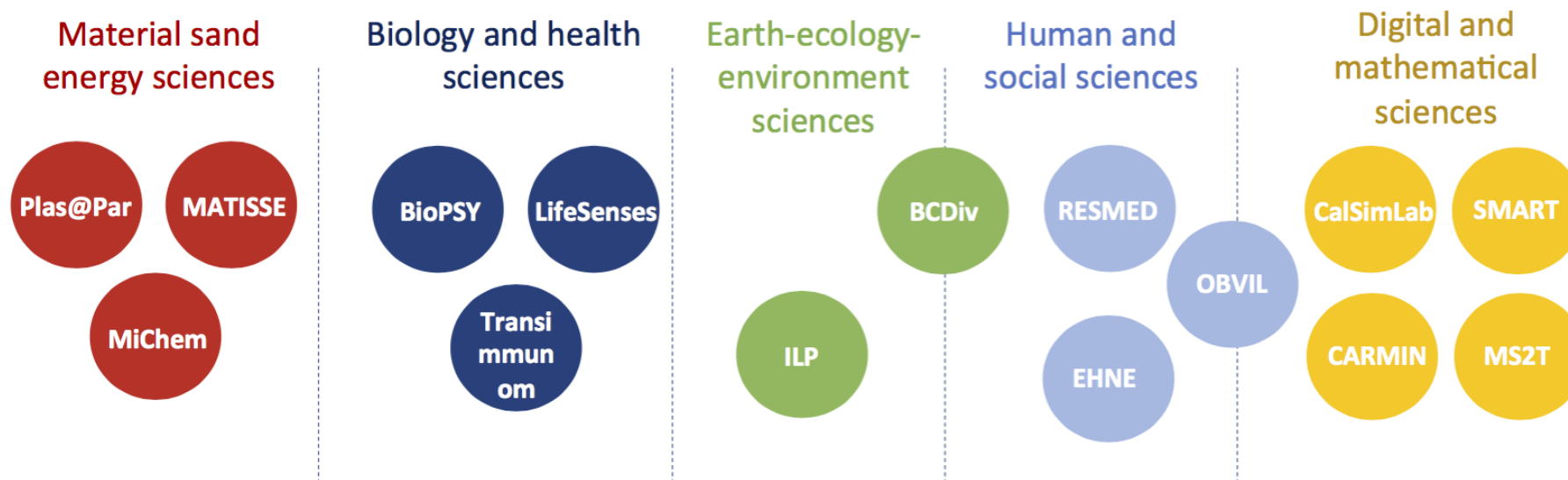
- un **constat**: la majorité des équipes de recherche des établissements partenaires de SU (comme d'ailleurs des autres centres académiques) n'est pas à même de répondre individuellement aux défis posés par le calcul et l'analyse de données;
- des **pistes**: pour éviter une nouvelle « **fracture numérique** », il est important d'accélérer
 - i) la **mise en réseau** des équipes impliquées dans le calcul, la production et l'analyse de données,
 - ii) la **structuration** et la **mutualisation** des **ressources** au sein d'une plateforme, et
 - iii) le déploiement de **programmes interdisciplinaires** autour du calcul et de l'analyse de données.
- seule une « **approche holistique** du calcul intensif » permettra de répondre aux **nouveaux défis** scientifiques (note du président du CNRS aux membres du COCIN)
- La stratégie préconisée est de s'appuyer sur des « pôles locaux associant infrastructures de calcul et de données, équipes de recherche inter et pluridisciplinaires autour de grands défis scientifiques en synergie avec les grands pôles universitaires de recherche et d'enseignement, les projets IDEX-IRT-LABEX, et les Régions ».

Contexte « Investissements d'Avenir »

- Sorbonne Universités
 - ComUE : regroupement de 11 universités et organismes de recherche



- plusieurs laboratoires d'excellence (labex) et équipements d'excellence (equipex)



- structuration de la recherche interdisciplinaire

Institut des sciences du calcul et des données

- créé en 2010 sur le modèle de centres de compétences universitaires européens et internationaux, ISCD regroupe **trois composantes** fondamentales :
 - i) un **centre de recherche**, d'expertise et de valorisation en calcul, analyse et visualisation de données,
 - ii) un **pôle de formation** et d'enseignement, et
 - iii) une **unité de service** et d'appui à la recherche;
- sa **mission principale** est de **motiver**, de **stimuler** et d'**organiser** la recherche multidisciplinaire en sciences du calcul et des données;
- l'**objectif** est de fournir des **ressources initiales** pour permettre à des initiatives scientifiques de voir le jour, d'**atténuer le risque** non-négligeable pris par des chercheurs qui s'aventurent dans l'exploration pluridisciplinaire ou à l'interface des disciplines, et de **favoriser** des formes de **dissémination des résultats** et de l'information innovantes;
- les **résultats obtenus** par cette initiative sont évalués par des **métriques classiques**: nombre de papiers écrits par des chercheurs et publiés, nombre de projets soumis en réponse à des appels à projets nationaux et européens.

Moyens

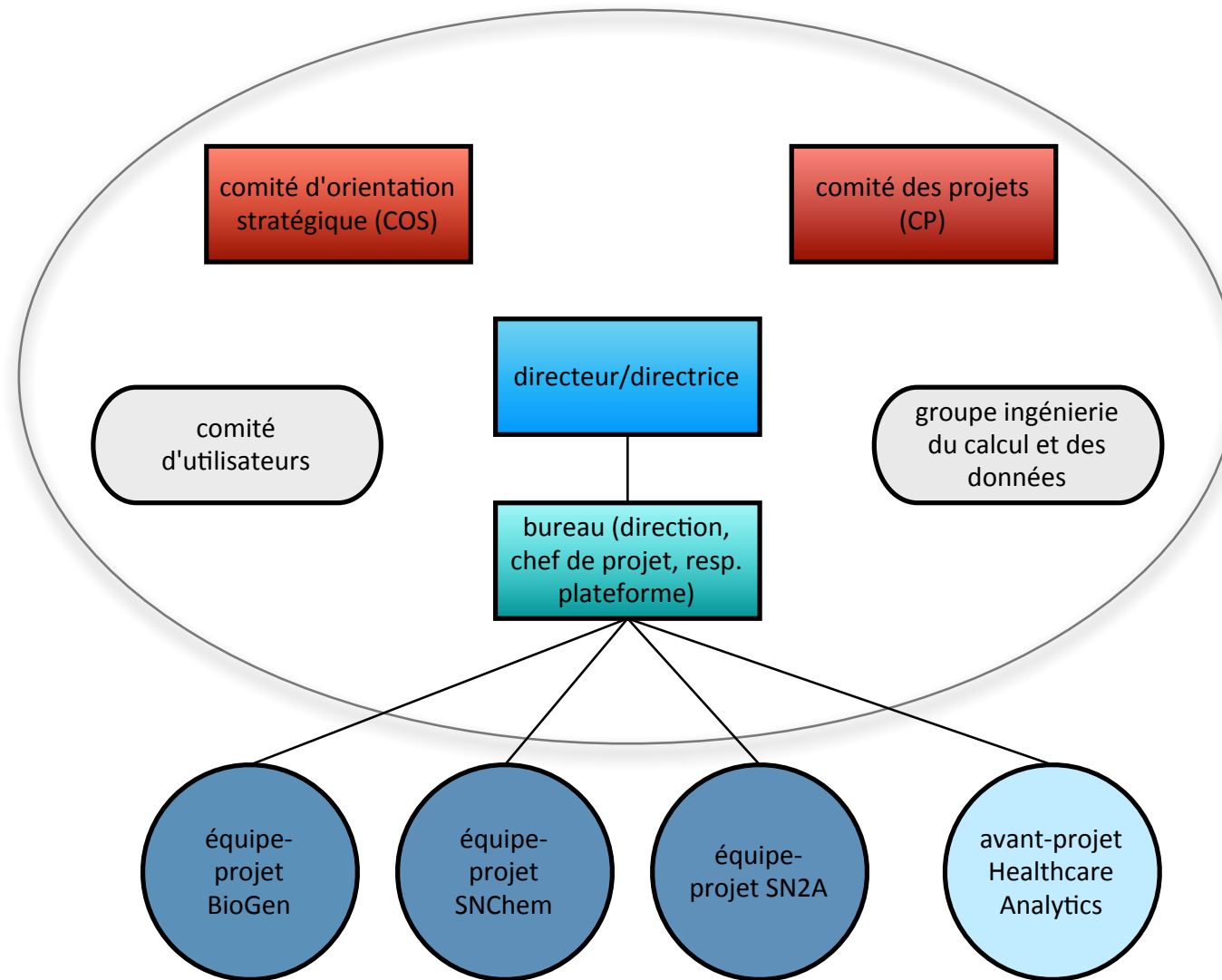
pour faciliter ces interactions, ISCD met en place les moyens suivants:

1. un [espace de coworking](#) pour héberger les chercheurs, ingénieurs, doctorants et étudiants impliqués dans des collaborations interdisciplinaires;
2. des [crédits d'animation scientifique](#) pour organiser des conférences, congrès, groupe de travail, écoles thématiques, ainsi que des sessions de réflexion dans lesquelles des chercheurs de différentes disciplines se réunissent autour d'un défi aux interfaces pour proposer des pistes et un programme de travail scientifique;
3. des [ressources humaines](#) (ingénieurs de recherche) et des [moyens de calcul et d'analyse](#) pour co-développer des programmes de calcul et d'analyse performants, prototyper des algorithmes et étudier leur scalabilité;
4. un [portail internet interactif](#) pour décrire les projets et les initiatives en cours, avec la liste des participants, ainsi que les principaux résultats de recherche;
5. la [diffusion](#) et la [promotion](#) des développements logiciels en mode "open-source".

Organisation de la recherche

Le **modèle d'organisation** de la recherche de l'institut est inspiré de celui mis en place à INRIA. Il repose sur la notion d'**équipe-projet pluridisciplinaire**, en tant qu'élément structurant.

- une **équipe-projet** est composée d'un ensemble de membres (chercheurs, ingénieurs et doctorants) rassemblés autour d'un « leader » scientifique qui établit des objectifs scientifiques et veille à leur mise en œuvre sur une thématique approuvée par le comité d'orientation stratégique (COS). Le programme de recherche est clairement défini et limité dans le temps (quatre ans);
- un **avant-projet** précède la constitution d'une équipe-projet pendant une phase de maturation qui a pour objectif de préciser les axes scientifiques et les thématiques de la future équipe et d'en identifier les membres permanents, pour la durée du contrat. Il dispose des ressources spécifiques de même nature que celles d'une équipe-projet (stages master, thèse, post-doctorat, support ingénieur de recherche).
- actuellement, l'institut héberge **3 équipes-projets** et **deux avant-projets**.



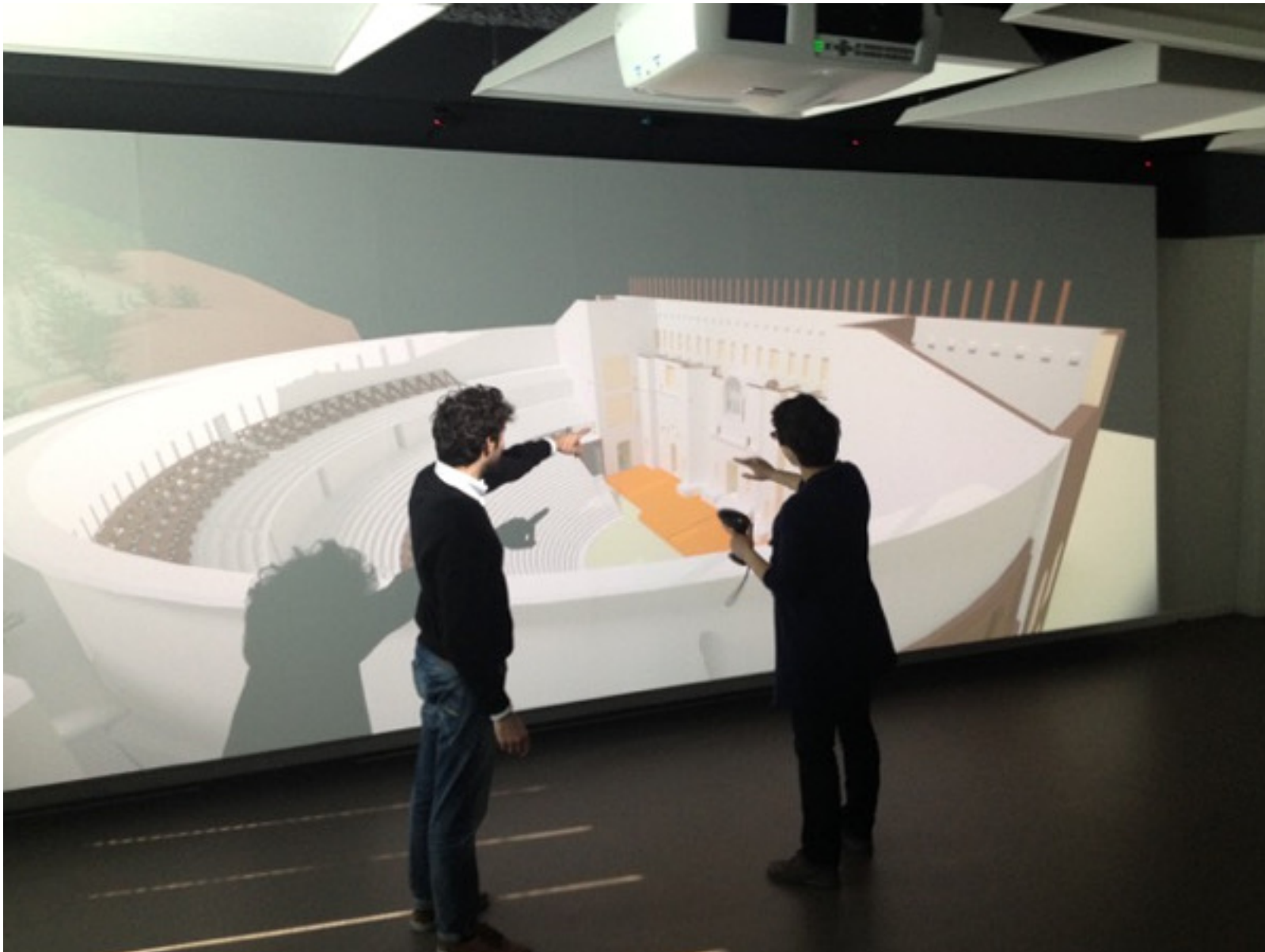
Organigramme de la gouvernance ISCD

Moyens calcul

- **Mesu-alpha**: SGI UV2000 (16Tb, 1024 cœurs)
 - 128 Intel Xeon E5-4650L 2.6Ghz (8 core)
 - 16 Tb RAM, 128 Go / proc
 - Numalink 6 12.5 Gb/s
 - 90 Tb storage
 - **19.5** Tflops
- **Mesu-beta**: cluster SGI ICE-XA
 - 168 Intel Xeon E5-2670v3 2.3 Ghz (12 core)
 - 128 Gb RAM / node
 - 4 graphic cards NVIDIA K5200
 - 200 Tb storage
 - **59.3** Tflops



Moyens visualisation



Réalisations première phase (2010-2016)

Plusieurs actions d'envergure ont été initiées et ont permis d'enclencher une dynamique de collaboration et de recherche interdisciplinaire sans précédent, avec des financements divers:

- PIA : **labex Calsimab** (5M€)/ **equipex equip@meso** (1.3M€) : 30 doctorants, post-doctorants ou IGR recrutés depuis 2013, invitations de chercheurs, 1 super-calculateur acquis pour la simulation intensive, école d'été « calcul et analyse aux interfaces »,
- **Sesame IdF RefICS** (720 k€) : plateforme de visualisation scientifique (2013-2017),
- **FEDER ICS@SU** (380 k€), budget plateforme UPMC (400k€) : contribution à l'acquisition d'un super- calculateur pour l'analyse de données (2016),
- **Idex SU** : SATS-SU (720k€), chaire thématique FaciLe (1,2M€), programme convergence NumerO : 2 titulaires senior, 3 doctorants et post-doctorants, 1 IGR, 2 stagiaires master, 1 assistante de projet
- **MI CNRS** (100k€/an, 3 ans) : 10 stages de bi-master, réalisation de courtes vidéos de promotion des métiers du calcul et de l'analyse de donnée (2016),
- **Mécénat EDF** : super-calculateur (ex-lvanoé) de 18 000 cœurs (280 Tflops).