



ICSD @ SU

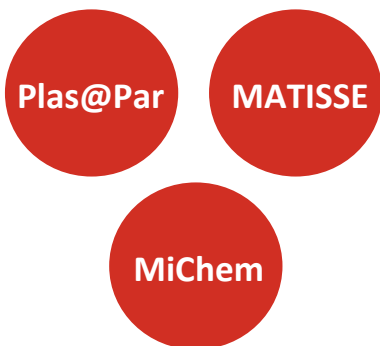
INSTITUT DU **C**ALCUL ET DE LA **S**CIENCE DES **D**ONNÉES

Sorbonne Universités

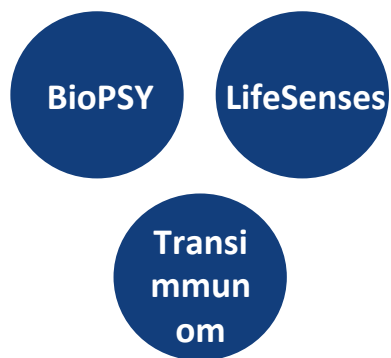
MEMBRES IDEX et LABEX



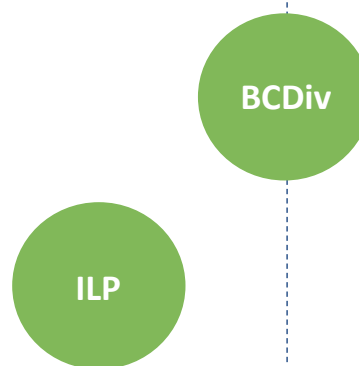
Energie, matière et univers



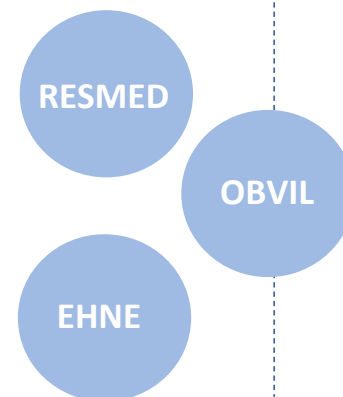
Biologie et santé



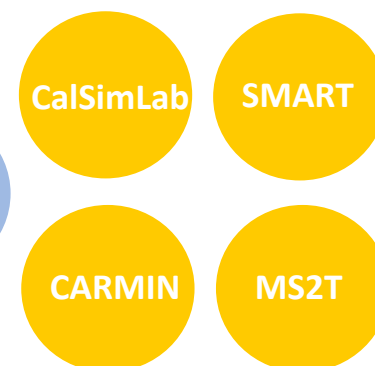
Terre vivante et environnement



Sciences humaines et sociales



Modélisation et ingénierie



Calcul et Science des Données

- deux domaines connaissent une évolution importante :
 - le **calcul scientifique**
 - la **science des données**
- ces domaines
 - sollicitent **informatique** et **mathématiques**
 - adressent des problématiques différentes
- recherche **pluridisciplinaire** impliquant de multiples compétences
- piliers de la science : théorie, expérimentation, simulation et analyse de données

Cadre Général

une **double ambition** pour ICSD

- renforcer la **primauté** de SU en calcul intensif, simulation et visualisation aux interfaces des disciplines
- développer de **nouvelles pratiques de recherche** en science des données
- former les “**data scientists**” capables d’analyser et d’exploiter les données massives et complexes
- irriguer de nouveaux **domaines d’applications**

Le projet est construit sur deux piliers :

- **excellence scientifique**
- **innovation pédagogique**



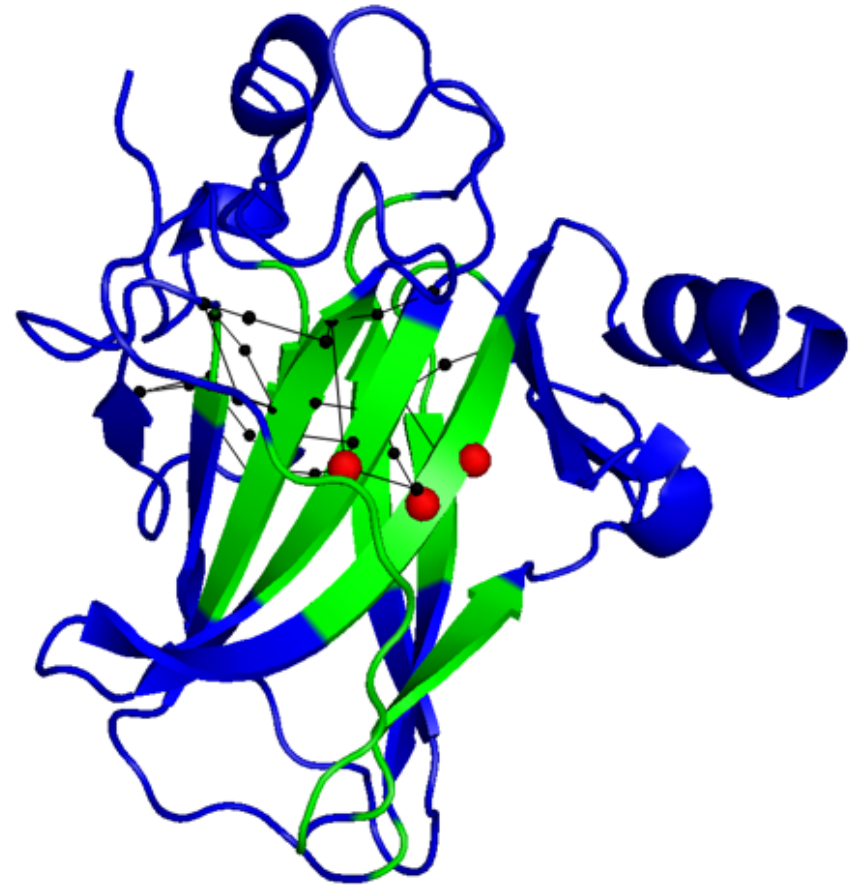
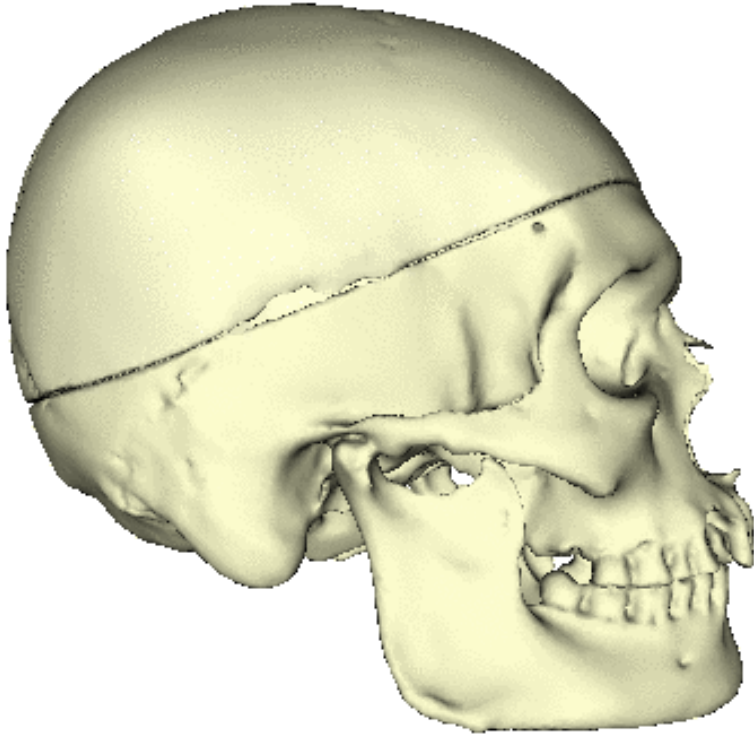
Plan Scientifique

- le calcul et la science des données progresseront à mesure de l'**avancée** du **front des connaissances** =>
 - développer les recherches les plus fondamentales et de
 - les coupler avec les activités directement liées aux domaines applicatifs disciplinaires
- deux défis majeurs à relever liés aux avancées
 - des **connaissances** et de la **formation**
 - des **technologies**
- atteindre les objectifs de l'évolution des domaines prioritaires (HPC, simulation numérique, traitement et transformation des données)

-> initiative ICS



Projets ICSD



axes thématiques: labex Calsimlab,
Chaire thématique, Convergence,
equipex,...

Enjeux Prioritaires

- nouvelles opportunités en **recherche, formation et valorisation.**
- les établissements doivent contribuer **collectivement**
 - à la collecte de données
 - à la formation au calcul, à l'analyse et à l'exploitation de ces données
 - à la recherche de pointe en calcul et en science des données
- un **défi** : « capacité à inter-opérer les domaines » et une **nécessité**
 - développer des infrastructures de HPC et de stockage massif
 - déployer à large échelle des outils d'analyse
 - piloter des systèmes complexes
 - coupler des approches par données et par modèles
 - concevoir de nouvelles méthodes mathématiques / applications
 - disposer de plateformes pour l'expérimentation et les tests

- gouvernance
 - comité exécutif (6 membres)
 - comité scientifique (12 membres)
 - identifie des actions de développement incitatives = projets transversaux, pluridisciplinaires et collaboratifs => mise en place d'une équipe-projet
- stratégie scientifique et actions incitatives
 - modèle repose autour de la notion d'**équipe-projet**
 - objectifs scientifiques + programme de recherche pour relever un défi
- moyens humains
 - enseignants-chercheurs, étudiants, professeurs invités, ingénieurs de recherche
 - contrats doctoraux

Etat d'avancement

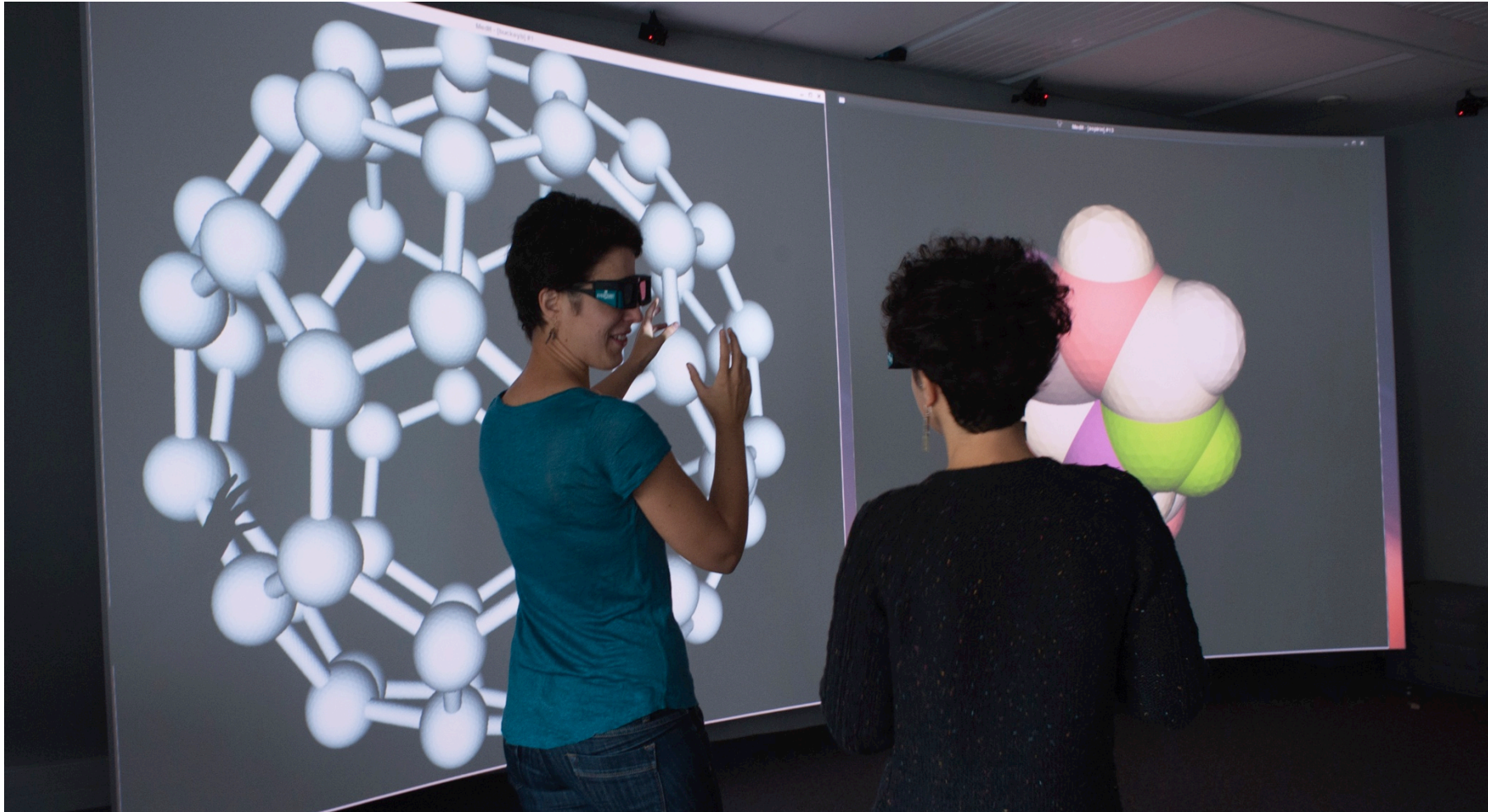
MOYENS CALCUL

- **Mesu-alpha:** SGI UV2000 (16Tb, 1024 cores)
 - 128 Intel Xeon E5-4650L 2.6Ghz (8 cores)
 - 16 Tb RAM, 128 Go / proc.
 - Numalink 6 12.5 Gb/s
 - 90 Tb storage
 - 19.05 Tflops
- **Mesu-beta:** cluster SGI ICE XA
 - 168 Intel Xeon E5-2670v3 2.3Ghz (12 cores)
 - 128 Gb RAM / node
 - 4 graphic card NVIDIA K5200
 - 200 Tb storage
 - 59.3 Tflops
- **Mesu-gamma:** cluster IBM iDataPlex
 - 20 Intel Xeon E5-2665 2.4Ghz (8 cores)
 - 256 Gb RAM / node
 - 4 graphics card NVIDIA Quadro 5000
 - 30 Tb storage
 - 3 Tflops



Etat d'avancement

MOYENS VISUALISATION



Lien avec le mésocentre IPSL

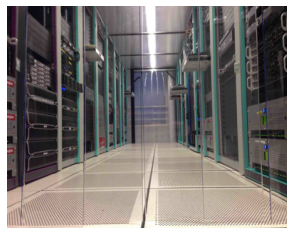


- communauté recherche sur le Climat (modélisations, observations, etc..)
- spécificité : très gros jeux de données,
- s'appuie sur deux structures situées à l'UPMC et à l'École Polytechnique
- Offre
 - stockage : environ 2.5 petaoctets
 - calcul : 2 grappes de calcul de 744 et 768 cœurs (28 nœuds), 10 Tflops
 - interconnexions locales : réseau 40 Gb/s, faible latence
 - interconnexion RENATER/RAP/SAFIR dédiée
 - logiciels : compilateurs, outils visualisation, gestion base de données,...
 - portails internationaux de redistribution de données

Interconnexion IPSL - ICSD



Données « Climat » > 2 Po



Institut du Calcul et de la Simulation
SORBONNE UNIVERSITÉS



> 3000 cœurs

Liaison 2x10 Gb/s



- besoin important pour la communauté climat : accéder au « calcul intensif » pour de très gros jeux de données
- définition d'un protocole commun d'accès aux données
- organisation système et ressources humaines