

GRENOBLE ALPES RECHERCHE
INFRASTRUCTURE DE
CALCUL INTENSIF
ET DE DONNÉES

L'UMS GRICAD

Un modèle organisationnel
original au service des
données et du calcul

Christian LENNE
Université Grenoble Alpes



2016 : Contexte grenoblois bouleversé

- Fusion des trois universités du site :
 - Joseph Fourier (Grenoble 1),
 - Pierre Mendès-France (Grenoble 2),
 - Stendhal (Grenoble 3)

Université Grenoble Alpes

- Regroupement de laboratoires de Maths/Info dans un même bâtiment :
 - LIG, LJK, VERIMAG
- Obtention d'un IDEX

Création de l'UMS GRICAD

- Création au 1^{er} Janvier 2016
- Enjeux :
 - **Calcul intensif**, traitement, diffusion et préservation des **données scientifiques** au cœur des activités de recherche.
 - Problématiques du calcul et du traitement intensif de plus en plus **indissociables** de l'exploitation et la valorisation des grandes masses de données.
 - **Contexte contraint** sur le plan énergétique, financier, humain.

Nécessité de repenser l'**urbanisation des infrastructures de calcul et de données** ainsi que l'**organisation des moyens humains** en synergie avec les laboratoires de recherche et les services communs.

CONTEXTE DE MISE EN PLACE

Contexte et historique du site

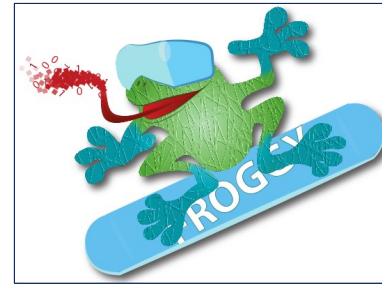
- **Projet CIMENT** autour du **calcul intensif** avec un modèle organisationnel distribué mixte très pertinent.
 - Une puissance de calcul **accessible à tous les personnels** des établissements de recherche académiques du site grenoblois
 - Un **réseau d'ingénieurs**, impliqués dans des équipes de recherche, deux ingénieurs dédiés
- **Projet SUMMER, plateforme mutualisée de stockage de données**, co-construit par des personnels de laboratoires et de services centraux
 - **Squelette de l'infrastructure** financé par l'UJF.
 - **Informaticiens (labos, services, composantes)** détachés pour une partie de leur temps sur l'exploitation de l'infrastructure de stockage.
- **Engagement politique** des membres des établissements qui s'est traduit par la volonté de **mutualiser et de rationaliser l'infrastructure datacentre** sur le site à l'horizon 2020.



- **Deux plateformes principales :**

- **Froggy** (calculs parallèles),

- **Luke** (calculs séquentiels, traitement massif)



- **CIGRI** : Ressources accessibles en **mode grille**

- Stockage **IRODS** permettant l'accès aux données réparties sur différents sites et sur des supports hétérogènes de façon transparente pour l'utilisateur.

- **Organisation en pôles** : un responsable technique et un responsable scientifique représentant les communautés de recherche.
- **Modèle humain** : un réseau d'ingénieurs, impliqués dans des équipes de recherche, deux ingénieurs dédiés.

SUMMER – Stockage Unifié Mutualisé Massif Evolutif et Réparti

- Etude en **2012** autour des besoins en stockage des laboratoires, composantes et services communs.
- Mise en place en 2013 d'une **infrastructure de stockage mutualisée**, pérenne et évolutive couvrant la majorité des besoins de stockage en s'adaptant le plus possible aux différents niveaux de service requis.
- Squelette de l'infrastructure financé par **l'UJF**, pour permettre ensuite aux entités de pouvoir acquérir de la volumétrie et un niveau de service associé.

- **Organisation**

- Comité Technique Stockage opérationnel : informaticiens (laboratoires, composantes, services communs) détachés d'une partie de leur temps pour exploiter l'infrastructure de stockage.



Datacentre grenoblois

- Volonté de la COMUE de **mutualiser et rationaliser** les infrastructures de salles informatiques.
- Vers un **datacentre de site** :
 - Infrastructure immobilière et technique visant à concentrer les équipements informatiques
 - Dans des salles sécurisés, redondantes, efficaces énergétiquement
 - Avec des infrastructures techniques spécifiques (alimentation électrique, refroidissement, accès au réseau)



ORGANISATION ET MISSIONS DE L'UMS GRICAD

Unité Mixte de Services GRICAD

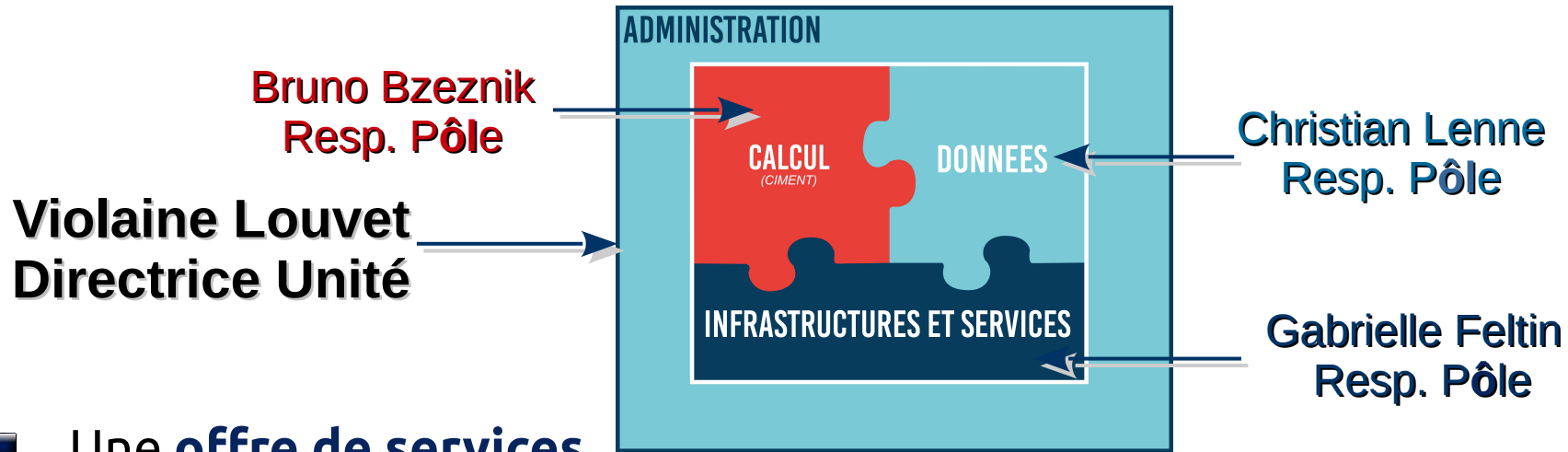


- **Missions autour du calcul intensif**
 - Gestion des plateformes de calcul et de traitement intensif et accompagnement des utilisateurs.
- **Missions d'infrastructures de base en co-pilotage avec les structures partenaires**
 - Gestion des salles des datacentres mutualisés du site grenoblois, hébergement, réseaux
 - **Missions d'expertises, de réseautage et de soutien**
 - Réseau de compétences.
 - Veille technologique et expérimentation.
 - Support au montage de projets nationaux ou européens sur la partie technique.
- **Missions autour des données**
 - Gestion des infrastructures mutualisées de stockage, de diffusion et de sauvegarde.
 - Support aux utilisateurs sur tout le cycle de vie de la donnée.
- **Missions d'infrastructures et de services**
 - Administration des services à valeur ajoutée pour la recherche : forge logicielle, bases de données...
- **Modèle organisationnel mixte**
 - Personnels rattachés à la structure et personnels des unités, services communs et composantes investis à temps partiel dans les projets.
 - Favorise la proximité avec les équipes de recherche, facteur de montée en compétence.



Organisation et offre de services

■ Organisation de GRICAD en pôle, avec un fonctionnement par projet



■ Une offre de services

- Soutenue par des Comités Techniques mixtes
- Cohérente et rationnelle avec les autres acteurs : DSI des établissements, SIMSU
 - Hébergement de serveurs
 - Machines virtuelles
 - Espace de stockage (SUMMER)
 - HPC, Traitement massif (CIMENT)

Mise en place du pôle données

- On part d'une feuille blanche
- Organisation d'un brainstorming avec structures diverses :
 - 6 personnes contactées, ~30 présentes
 - 4 axes de travail ont émergé :
 - Confidentialité, Innovation juridique, sécurité
 - Qualité, Traçabilité, Cycle de vie
 - Traitements / Portail
 - Liens entre briques d'infra →Projet Prospectives autour de la virtualisation
- Animation des groupes par membres extérieurs à GRICAD

Groupes plus orientés vers le Pôle données

Axe 1 : Confidentialité, Innovation juridique, sécurité

- ✓ Acquisition de données
- ✓ Sécurisation d'accès
- ✓ Anonymisation
- ✓ Evolutions juridiques
- Open data

Axe 2 : Qualité, Traçabilité, Cycle de vie

- ✓ Nettoyage de données
- ✓ Traçabilité
- ✓ Reproductibilité de traitements
- ✓ Plan de gestion de données

Axe 3 : Traitements / Portail

- ✓ Plates-formes expérimentales
- ✓ Environnement de tests
- ✓ Techno BDD (SQL, NoSQL, SIG)
- ✓ Outils statistiques
- ✓ Spécificités d'enquêtes
- ✓ Traitement de données massives du WEB
- ✓ Requêtage, présentation
- Valorisation des données

Axe 4 : Liens entre briques d'infra.

- ✓ Optimisation des liens entre briques CIMENT, SUMMER, VM
- ✓ Définition d'architectures nouvelles
- ✓ Veille et test technologies émergentes
- ✓ ...

Organisation de l'aspect formation à envisager

RETOUR SUR LES PREMIERS MOIS

3 participations à la soumission de projet IDEX

- **CALCUL :**
 - CDP UGA 2.0 : plateforme intégrative
- **Données :**
 - Institut de la donnée (CDP Grenoble Data Institute)
 - Suivi d'évolution de territoires et de la flore (CDP Trajectories)

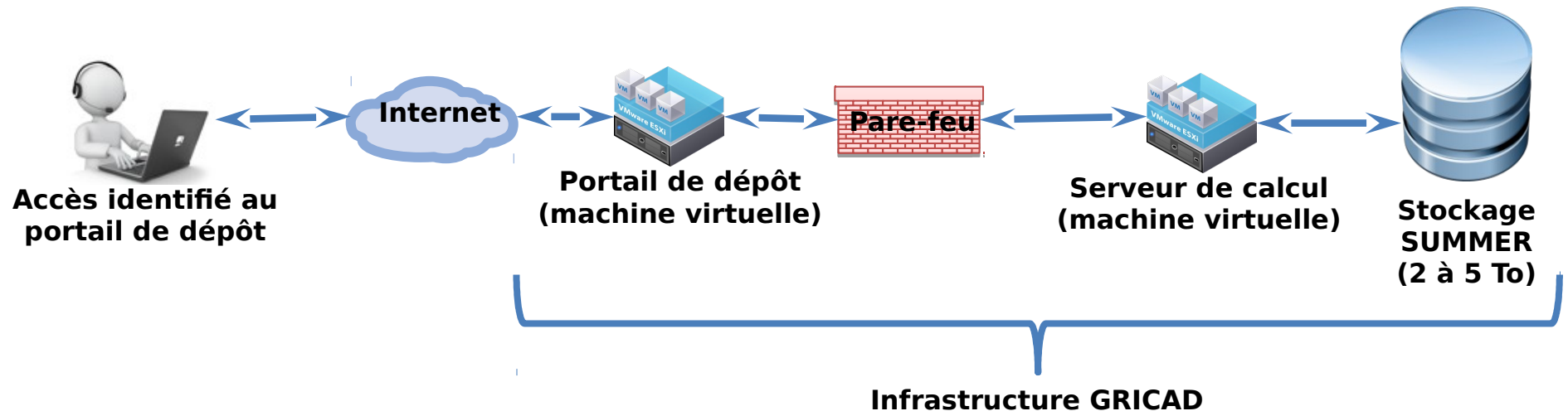
Participation au projet européen EOSC (European Open Science Cloud)

- WP Interoperability
- « Science Demonstrator » basé sur l'utilisation de notebooks pour l'accès transparent aux ressources de calcul et de données

Réponse à des demandes d'infras

- Projets en Santé
 - H2020 : Portail suivi traitement AVC
 - Neuro : Portail recherche sur images IRM
 - EpiMed : Portail dédié à l'accès de données de type « OMICS »
- Projet ANR
 - E-Fran : portail de collecte et de traitement de « traces » d'apprentissage de la lecture

Architecture demandées



Conseils et assistance

- Evolution méthode/infrastructure suivi épidémiologique équipes Laboratoire de Biologie Moléculaire
- Optimisation de code pour parallélisation d'algorithme en traitement de données d'économie.
- Méthologie liée à de l'étude épidémiologique sur des maladies professionnelles : application de techniques de Syst. Info. Géographique
- CdC / Architecture entrepôt de données de santé pour la recherche et le soin

- Travail sur des architectures de virtualisation adaptées au traitement de données « Big Data »
- Mise en place d'un démonstrateur déployable pour un portail d'accès et de traitement de données structurées et non structurées (entrepôt de données en santé)

Merci pour votre attention

Questions ?

Site Web : <https://gricad.univ-grenoble-alpes.fr>

E-Mail :

Direction : [Violaine.Louvet At univ-grenoble-alpes.fr](mailto:Violaine.Louvet@univ-grenoble-alpes.fr)

Calcul : [Bruno.Bzeznik At univ-grenoble-alpes.fr](mailto:Bruno.Bzeznik@univ-grenoble-alpes.fr)

Données : [Christian.Lenne At univ-grenoble-alpes.fr](mailto:Christian.Lenne@univ-grenoble-alpes.fr)

Infra : [Gabrielle.Feltin At univ-grenoble-alpes.fr](mailto:Gabrielle.Feltin@univ-grenoble-alpes.fr)