



Inserm



La science pour la santé _____
_____ From science to health

SCaaS – Scientific Computing as a Service Vers une offre « informatique scientifique » à l’Inserm

Inserm DSI/SSDUN, domaine
Informatique Scientifique

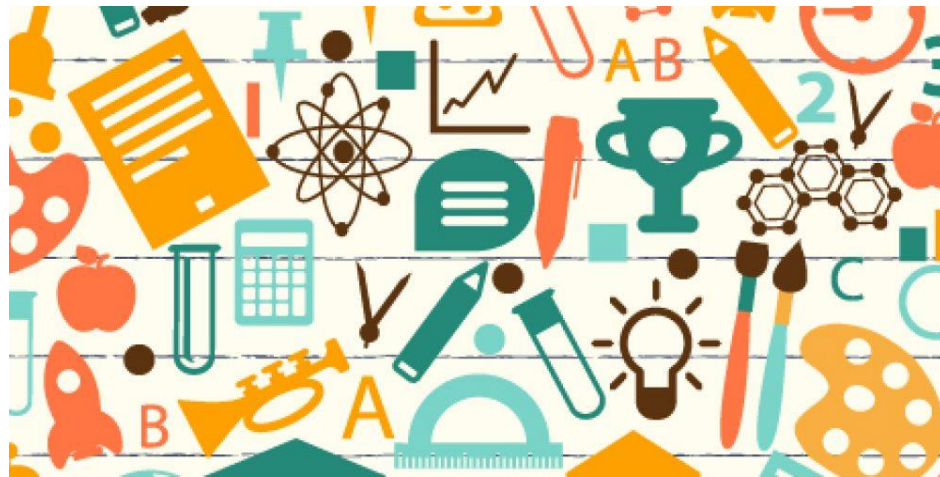
JCAD2021 – Décembre 2021

Quelques mots sur l'Inserm

- Institut dédié à la recherche biologique, médicale et à la santé humaine
- Organisation et gouvernance
 - Cotutelle MESRI et Ministère de la Santé
 - 9 Instituts Thématiques (axes scientifiques)
 - 13 délégations régionales (axes administratifs)
 - Des départements transversaux, dont le DSI
- Structures de Recherche
 - 281 Unités de Recherche (toutes mixtes)
 - 36 Centres d'Investigation Cliniques
 - 34 Unités de Service

Le calcul et les données à l'Inserm

- Une grande hétérogénéité
 - Des types de données et des formats
 - Des volumétries et des profils d'utilisation
 - Des outils et des workflows
 - Des solutions en place



Constat à l'échelle de l'Institut (1)

- **Stockage et volumétries**

- 50%+ des unités ont entre 10 et 100To de données actives, et autant de données historiques
- 68% des unités produisent moins de 10To/an
- Evolution variable
 - de stable à fortement exponentielle.
 - Accroissement massif anticipé pour les données liées au séquençage (plusieurs Po/an)

- **Données nécessitant une protection**

- 50% des unités manipulent des données à caractère personnel
- 75% des unités manipulent des données considérées comme « sensibles »
- Ces données sont très majoritairement stockées en interne et/ou dans les CHU

Constat à l'échelle de l'Institut (2)

- **Infrastructures**
 - Solutions locales fortement majoritaires
 - Centres régionaux : utilisation croissante
 - Centres nationaux et infrastructures nationales : utilisation encore marginale

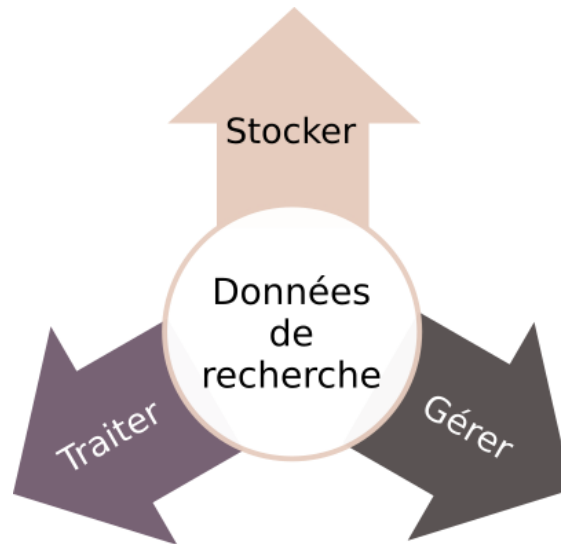
- **Solutions de calcul parfois inadaptées**
 - Problématiques de connexion réseau
 - Gros transferts de données amont/aval
 - Workflows/outils spécifiques
 - Manque de technicité/compétence

Besoins remontés par nos unités de recherche

- **Besoins en termes de gestion de la donnée:**
 - Stockage classique
 - Stockage sécurisé
 - Sauvegarde et Archivage
 - Partage
 - Gestion de données ouvertes (publications/DOI)
- **Besoin en termes de traitement**
 - Capacité de calcul
 - Solutions de partage logiciel
 - Flexibilité de l'environnement
- **Besoins en termes d'accompagnement**
 - Support de proximité et de qualité
 - Formations techniques

Le projet SCaaS : Scientific Computing as a Service

Des services centrés sur la donnée de recherche



Des services interconnectés :

- Humainement : expertise et accompagnement
- Techniquement : accès unifié aux services



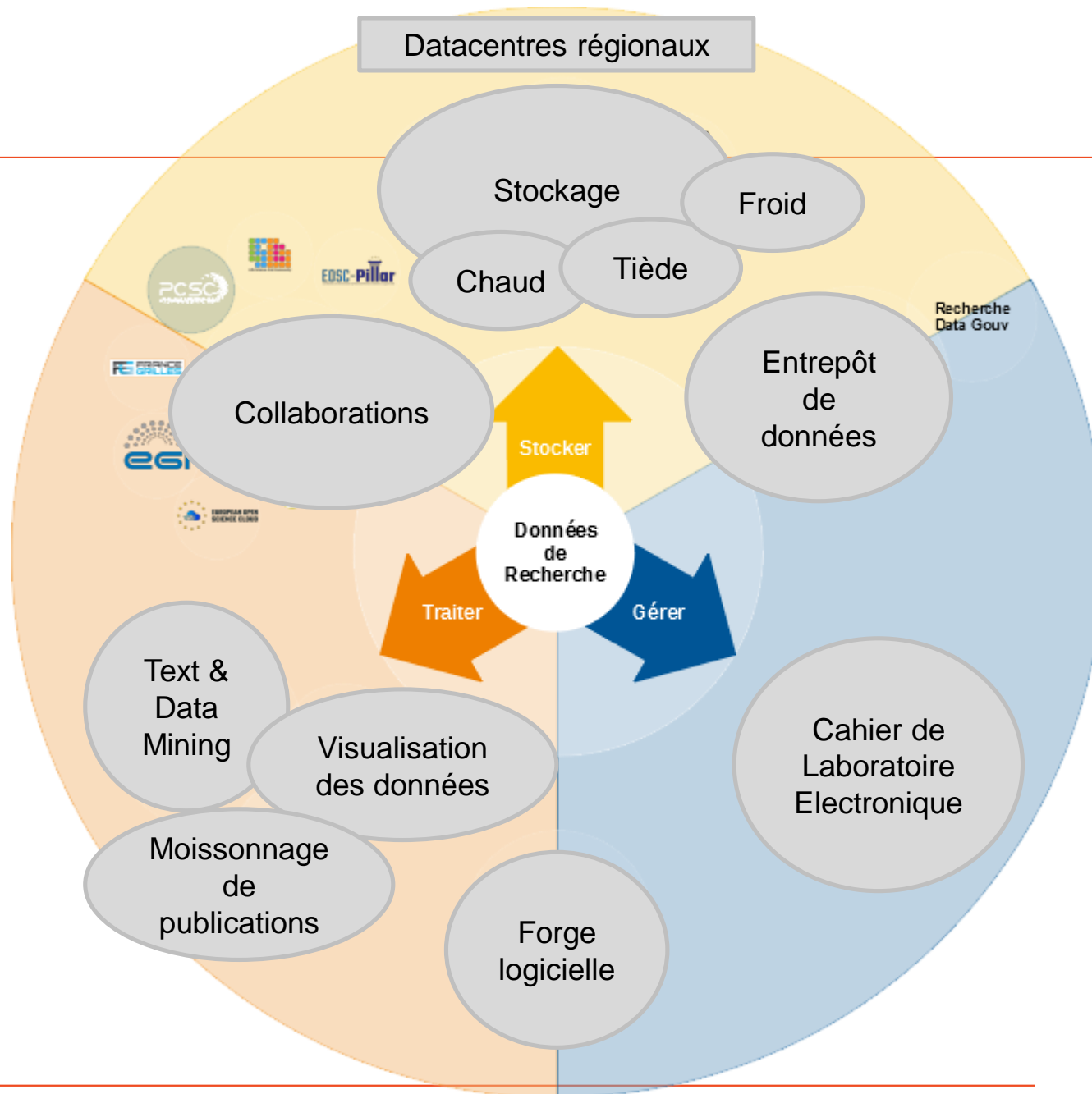
Objectifs visés

Répondre aux besoins des chercheurs

- Proposer des offres de stockage et de gestion de données
- Proposer et orienter vers des moyens de calcul
- Assurer un service privilégié

Répondre aux exigences de l'Inserm

- S'insérer dans le paysage européen
- Être présent sur le terrain national
- Mutualiser les moyens et harmoniser les ressources
- Répondre aux nouvelles exigences de science ouverte



Où en sommes nous ?

- **Réalisé**
 - Cahier de Laboratoire Electronique (CLE)
 - Support mutualisé (France Grilles, EGI)
 - Collaboration établies (PCSC, EOSC)
 - Moissonnage de publications scientifiques
- **En cours**
 - Offre de stockage (chaud, tiède, froid)
 - Distribution en datacentres régionaux
 - Entrepôt de données
- **En projet**
 - Forge logicielle
 - Structuration d'une cellule d'accompagnement

Pour en savoir plus...

- Consultez le poster dédié :
 - (lien depuis le site des JCAD2021)
- Ecrivez-nous :
 - ssdun-domis.dsi@inserm.fr
- Parlez-nous !
 - Quand on pourra enfin se retrouver en vrai, enlever nos masques et se raconter des choses autour d'un verre dans une session poster conviviale, parce que bon, là, c'est quand même un peu compliqué d'échanger même si c'est mieux que rien. OK, j'arrête de râler. Mais je teste quand même votre vision pour savoir jusqu'où vous pouvez encore lire les imbécilités que j'écris

Inserm Le service pour la santé. Pour améliorer la santé.

SCaaS – Scientific Computing as a Service

Vers une offre de service « Informatique Scientifique » à l'Inserm

Gilles MATHEU^{1,2}, Dominique PIGEON¹, David SALAS^{1,2}, Yvona SAMAA^{1,2}, Fanny BRIZZI^{1,2}, Jean Renaud MURLIN^{1,2}, Julo MARTINS¹

¹ DSI Inserm - Services Solutions et Développement des Usages Numériques (SSDUN)
² Equipe "Informatique Scientifique" du DSI de l'Inserm - ssdun-domis.dsi@inserm.fr
³ Réseau de Recherche et d'Innovation - RRI
⁴ DSI Inserm - Mission Infrastructures

Présentation

En 2019, le Domaine Informatique Scientifique de l'Inserm (DomIS) s'est lancé dans une réflexion sur la mise en place d'une offre « informatique scientifique » à destination de la communauté Inserm (Unités de Recherche ou de Service, équipes, chercheurs, projets...).

Démarche

- Collecte et analyse des besoins en stockage, partage, calcul...
- Étude de l'offre existante

Objectifs

- Augmenter les ressources disponibles (stockage, calcul)
- Assurer un service privilégié
- Contribuer aux offres en place
- Soutenir la collaboration inter-instituts

Offre de services Informatique Scientifique

Logiciel Openstack - Test - Stockage

Stockage Primaire

- NAS
- Data Centre
- Répertoire

Stockage Tiède

Tiering stockage Objet

Stockage Froid

Réplication vers AWS

Entrepôt de Données

- Gestion
- Partage
- Publication
- Principe FAIR

Recherche Data Gov

Calvier de Laboratoire Electronique

- Gestion au quotidien
- Collaboration
- Mise en qualité

Forge Logicielle

- Faciliter la gestion
- Partage
- Publication codes

Messagerie publications

Visualisation des données

- Text Data Mining
- Enrichissement des métriques
- Classification
- Ontologies

Collaborations

- Cas d'étude de services
- Support calcul
- Projets conjoints

Données de Recherche

Stockage

Traiter

Gérer

En service
En développement
En projet

Retours d'expérience

- Difficulté à établir un modèle économique "pratique" pour le stockage
- Importance du réseau utilisateurs (CLE)
- Bénéfices de la démarche "allians chercheurs" (ED)stockage)
- Intérêt de l'implication multi-institutionnelle (PCSC)
- Intérêt de la transversalité (TDM)

Perspectives

- Développer un lien entre les services
- une cohérence/cohesion entre l'entrepôt, le CLE et la forge logicielle
- une offre de stockage sur laquelle les services ci-dessus pourraient s'appuyer
- un lien entre les services "données de la recherche", les données administratives Inserm et le monde des publications scientifiques via le TDM

Visualisation: Nicolas Lécuyer et l'équipe de DataScience sur datascience.inserm.fr