

Gestion et ouverture des données de recherche à l'Inserm

Gilles MATHIEU, Dominique PIGEON, Daniel
SALAS, Marwa BELHAJ-SALEM, Nebras
GHARBI Jean-Renaud MURLIN, Sylvain
GERARD

Inserm DSI

101 rue de Tolbiac, 75013 Paris

Plan

- **Intro : l'Inserm et ses données de recherche**
 - Quelques infos sur l'Inserm
 - Données de recherche à l'Inserm : cadre stratégique
- **Projets lancés**
 - Modèle de Plan de Gestion de Données
 - Cellule d'accompagnement à la gestion des données
 - Offre de stockage massif distribuée
 - Entrepôt de données Institutionnel
- **Coordination des actions**
 - Gouvernance institutionnelle de la donnée
 - The "*big picture*" : vers un environnement numérique du chercheur

Quelques mots sur l'Inserm

- **Institut dédié à la recherche biologique, médicale et à la santé humaine**
- **Organisation et gouvernance**
 - Cotutelle MESR 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

Les données de recherche 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
- **Réflexions en constante évolution**
 - Présentation et poster aux JCAD2021
 - <https://doi.org/10.5281/zenodo.7148897>

Inserm
Le service pour la santé
SCaaS – Scientific Computing as a Service
Vers une offre de service « Informatique Scientifique » à l'Inserm

Gilles MATHEU¹, Dominique PIGEON², David SALAS³, Yvona SANAIA⁴, Fabry BRIZZI⁵, Jean-Renaud BURLE⁶, Jody MARTIN⁷
 (1) DSI Kerley - Service Système et Développement des Usages Numériques (SSDUN)
 (2) DSI Kerley - Service Système et Développement des Usages Numériques (SSDUN)
 (3) DSI Kerley - Service Système et Développement des Usages Numériques (SSDUN)
 (4) DSI Kerley - Service Système et Développement des Usages Numériques (SSDUN)
 (5) DSI Kerley - Service Système et Développement des Usages Numériques (SSDUN)
 (6) DSI Kerley - Service Système et Développement des Usages Numériques (SSDUN)
 (7) DSI Kerley - Service Système et Développement des Usages Numériques (SSDUN)

Présentation

En 2016, le Domaine Informatique Scientifique de l'Inserm (DISIS) 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...
 • Etude de l'offre existante

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

Offre de services Informatique Scientifique

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 "labors chercheurs" (Echocollage)
- Intérêt de l'implication multi-institutionnelle (PCSC)
- Intérêt de la transversalité (TDM)

Perspectives

Développer un lien entre les services

- une collaboration entre l'entité, la CLE et la forge logicielle
- une offre de stockage via 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

Cadre stratégique

Contrat d'Objectifs, de Moyens et de Performances entre l'Inserm et l'Etat 2021-2025

- Objectif 8 « *Renforcer l'accès ouvert aux données de recherche et aux publications de l'Inserm* »
 - Action 46 : mettre en place une gouvernance institutionnelle de la donnée
 - Action 47 : créer une structure d'accompagnement
 - Action 48 : concevoir un environnement numérique complet au plus près du cycle de vie de la donnée.
- Objectif 9 « *Mettre en œuvre une politique globale de gestion des données* »
 - Action 52 : Héberger les données de recherche dans les datacentres régionaux labellisés
 - Action 53 : Offrir du stockage massif de données
 - Action 54 : Mettre en place des bulles sécurisées

<https://fr.calameo.com/read/005154450d78f0e605668>

Modèle de Plan de Gestion de Données

- **Modèle construit par un groupe de travail transversal**
 - Inspiré de modèles existants (ANR et Institut Pasteur principalement)
 - Enrichi pour les problématiques liées aux sciences de la santé
- **Mis en ligne sur DMP OPIDoR**
 - Mai 2022 : version 1
 - Juillet 2022 : version 2 + modèle en anglais
- **Evolutions et perspectives**
 - Etudier la pistes de modèles disciplinaires
 - Modèles DMP structures
 - Evolution vers un modèle structuré

https://dmp.opidor.fr/public_templates?page=1&search=inserm

Accompagnement à la gestion des données (1)

- **Sujet d'actualité**
 - Besoin largement identifié
 - Volonté politique
 - En lien avec de nombreux projets de terrain
- **Besoin de transversalité**
 - Des compétences diverses et réparties dans plusieurs structures
 - Un réseau de terrain à mobiliser
 - Un lien à créer ou consolider avec les acteurs locaux
- **Quelques freins/risques identifiés**
 - Pas sûr que les chercheurs viennent d'eux-mêmes chercher l'info où elle est
 - Risque de rajouter une couche sur ce qui existe déjà

Accompagnement à la gestion des données (2)

- **Phase 1 : Construction organisationnelle**
 - Construction du groupe
 - Définition et cadrage (comitologie, gouvernance, périmètre, ressources...)
- **Phase 2 : Mise en œuvre pratique**
 - Mise en place des outils pratiques et construction du contenu
 - Définition du plan de communication et validation du calendrier
- **Phase 3 : Lancement**
 - Ouverture du service selon le calendrier
 - Structuration du réseau de terrain avec partenaires (universités, sites...)
 - Contribution Inserm aux « Ateliers de la donnée » du MESR
- **Phase 4 : Mode croisière et amélioration continue**
 - Opération du service, veille documentaire
 - Suivi de terrain, opportunités d'extension à d'autres sites/régions
 - Formation et auto-formation continue

Accompagnement à la gestion des données (3)

- **Mise en place d'un GT transversal (mai-décembre 2022)**
- **Travaux réalisés**
 - Recensement de l'existant (interne et externe) : viser la mutualisation et la coopération
 - Étude du besoin (retours d'expérience et enquête directive)
 - Définition des étapes de mise en place
- **Travaux à venir**
 - Définition d'une offre de service v1
 - Identification des compétences et ressources nécessaires
 - Validation du plan de mise en œuvre

Offre de stockage massif (1)

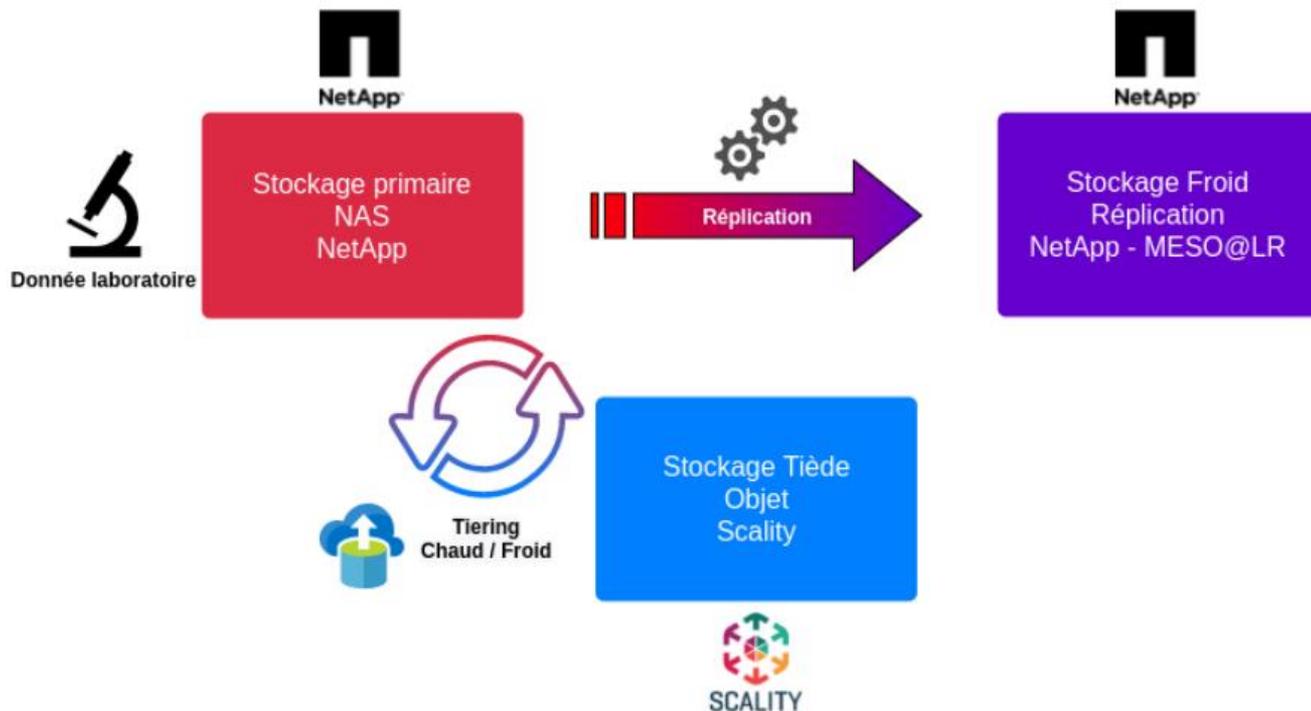
- **Objectifs**

- Standardiser et unifier les moyens de stockage institutionnels
- Offrir une alternative à l'acquisition propre

- **Modèle économique étudié**

- Coût fixe au To en fonction du niveau de service
- « Subvention » Inserm intégrée au tarif
- Refacturation du service aux laboratoires par le DSI
- Permettre le financement sur projet et récurrent

Offre de stockage massif (2) : topologie

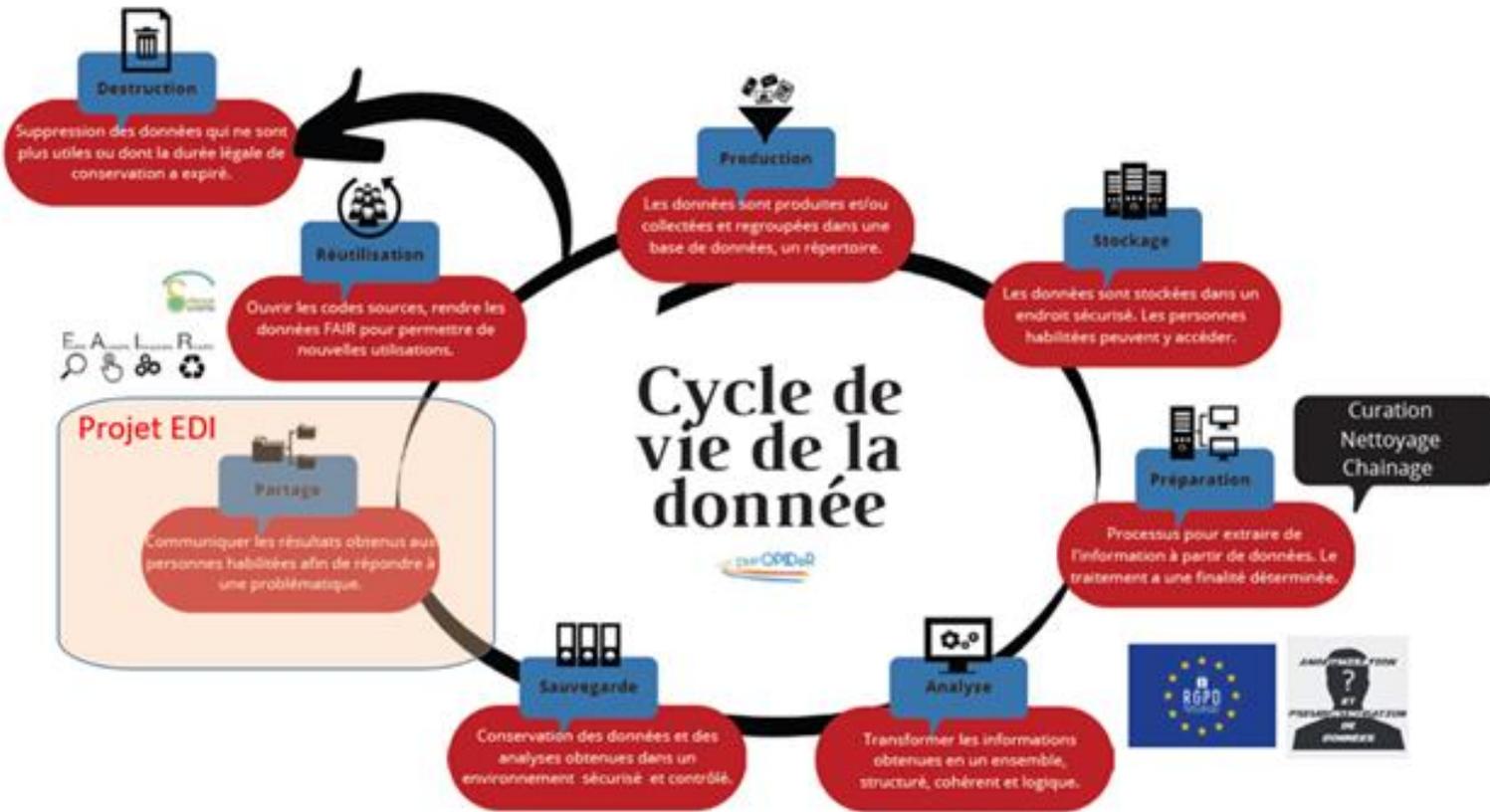


Entrepôt de Données Inserm : EDI (1)

- Un entrepôt de données à l'Inserm
 - Un entrepôt pour permettre le dépôt et la publication de l'ensemble des données d'un projet de recherche



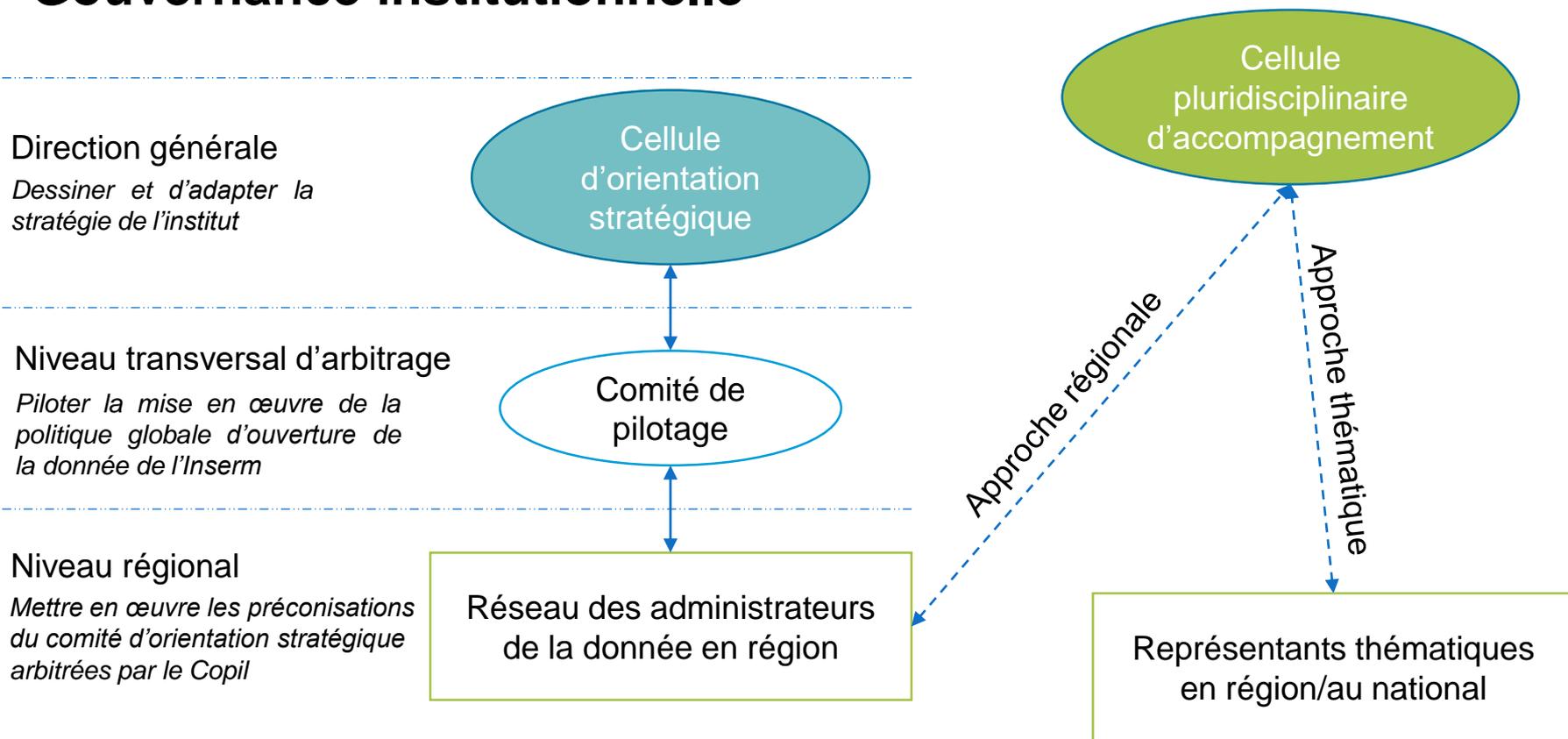
Entrepôt de Données Inserm : EDI (2)



Entrepôt de Données Inserm : EDI (3)

- **Fruit d'une longue réflexion**
 - POC avec data.inra.fr (2018-2019)
 - Construction du plan stratégique de l'Inserm (2019-2020)
 - Pré-Etude (2021)
- **Mise en oeuvre**
 - Entrepôt institutionnel au sein de la plateforme nationale **Recherche Data Gouv**
 - Cadrage fonctionnel en cours de validation
- **Prochaines étapes**
 - Recettes fonctionnelles (automne 2022)
 - Phase pilote avec un nombre restreints d'utilisateurs (fin 2022-début 2023)
 - Ouverture en production (mi-2023)

Gouvernance institutionnelle



The Big Picture

Gouvernance



Environnement numérique complet

Processus métier

Niveau fonctionnel

Services et Applications

Niveau technique

