



Données de santé au CEA: hébergement, services numériques et partenariats

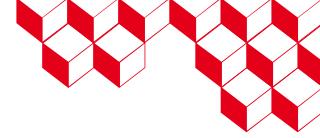


Assises Nationales des DOnnées Recherche

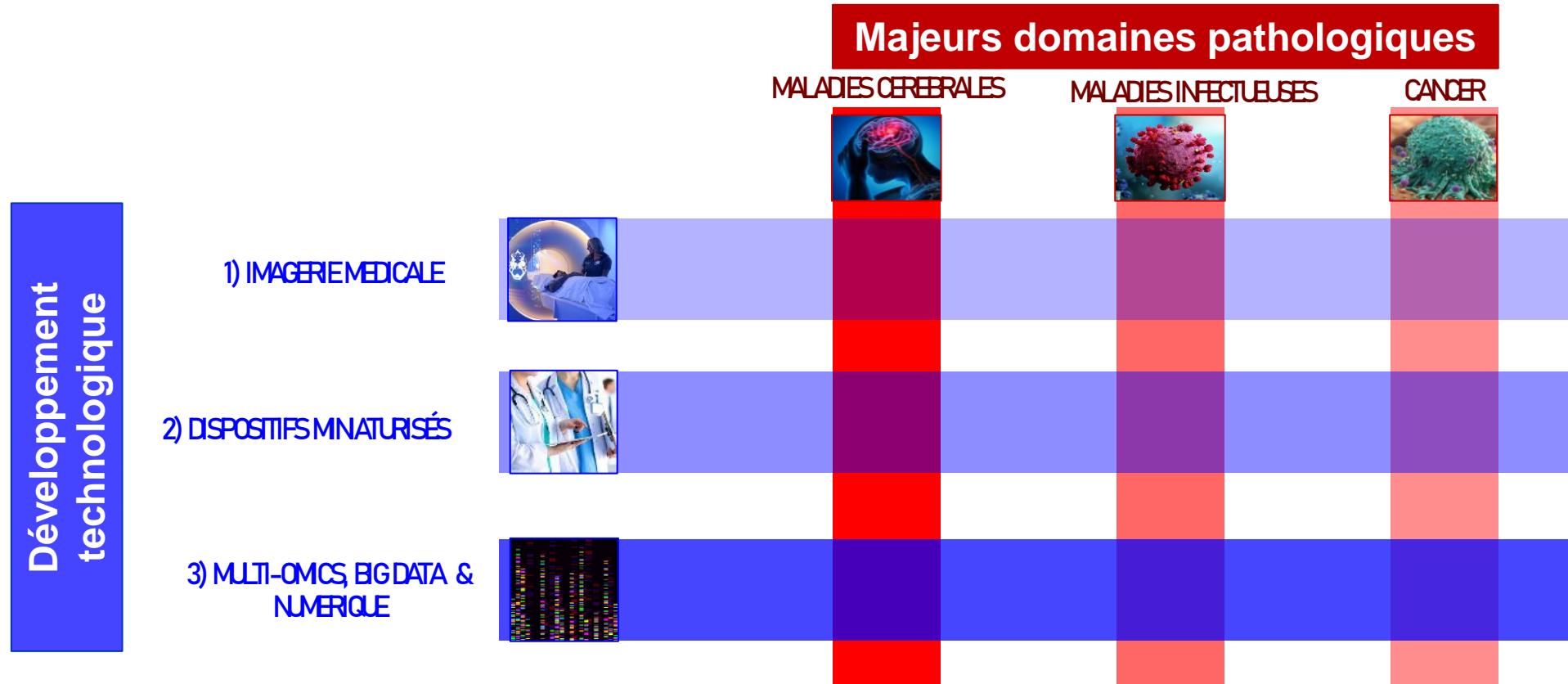
1er Décembre 2025 – ENS Paris-Saclay

yves.vandenbrouck@cea.fr

**Maison de la Simulation
Direction de la Recherche Fondamentale**



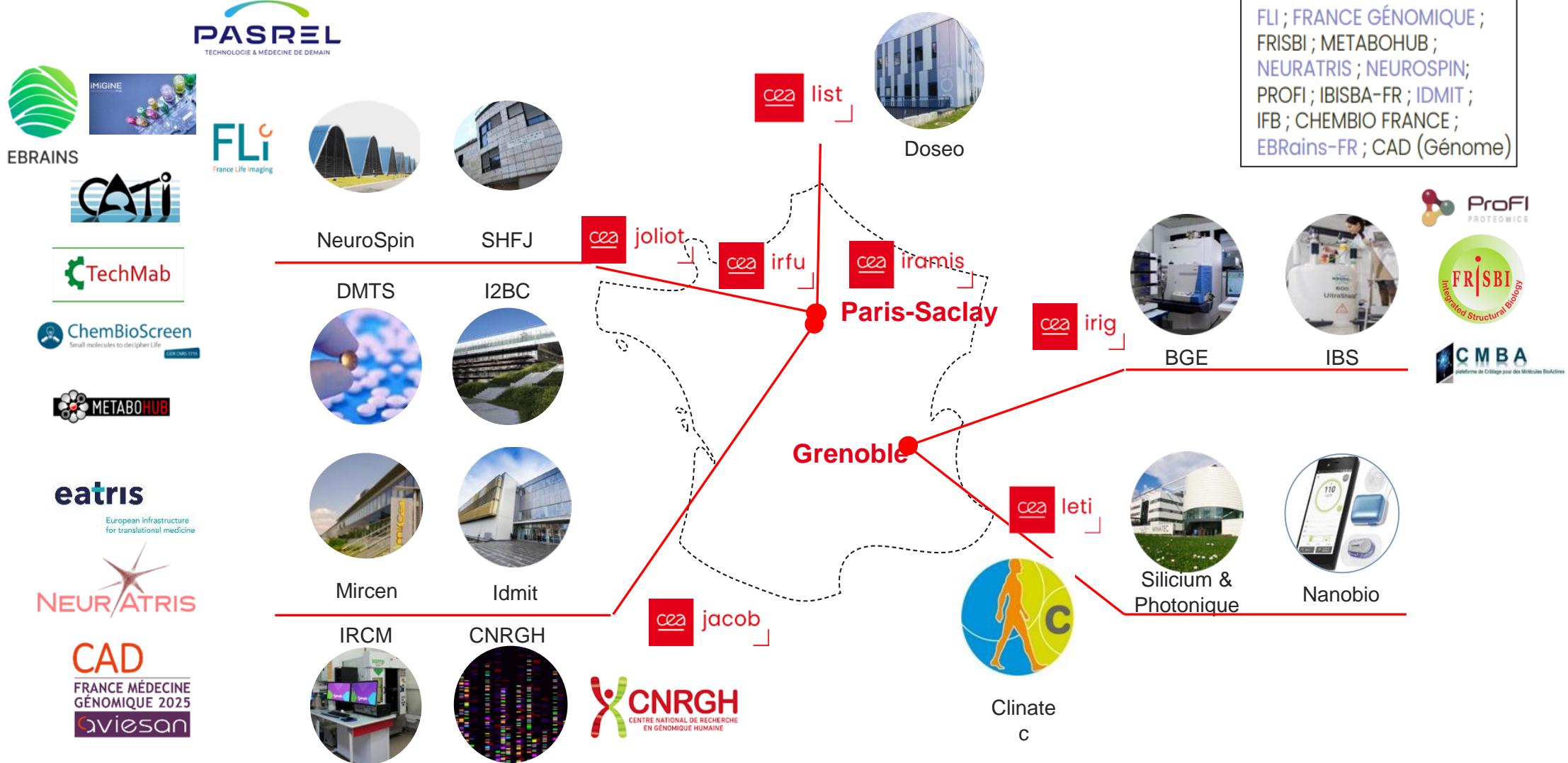
Un positionnement à la croisée de la technologie et de la médecine



- >1200 personnes dédiées dont ~450 PhD/Postdocs
- Près d'un millier de publications par an
- > 120 projets européens depuis 2016
- France 2030: 2 IHU (ICE, Sepsis), 6 PEPR (Biothérapies, Med-OOC, Santé Numérique, MIE, ProPSY, SAMS), PIIEC (Med4Cure)
- ~14 start-ups créées dans le domaine de la santé depuis 10 ans – partenariats industriels

Production de données au CEA

- 13 infrastructures nationales Biologie/Santé (6 co-pilotage CEA) et plateformes structurantes
- Des données produites pour extraire de nouvelles connaissances et alimenter modèles d'IA
- Des plateformes pour valider expérimentalement les prédictions issues de l'IA





Données de santé : développement d'une offre de services numériques pour la recherche

- **Production et gestion des données pour la recherche en santé : de qualité, accessibles, partageables et interopérables** (imagerie, génétique, « omiques », phénotypage multimodal,...) : ≠ échelles, modalités, volumétrie, informativité
- **Disposer d'une infrastructure numérique** pour maîtriser le **cyclde vie des données**, de l'hébergement jusqu'à l'exploitation, satisfaisant aux **exigences réglementaires pour l'utilisation secondaire des données**

3 infrastructures numériques en santé en cours au CEA:

TGCC

1. Partenariat CAD (Collecteur Analyseur de Données) du PFMG 2025
2. N4Brain (Numeric for Brain Research) avec CEA, INRIA, INSERM
3. EBRAINS 2.0 – E.U. (ESFRI)

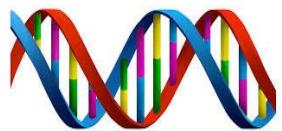
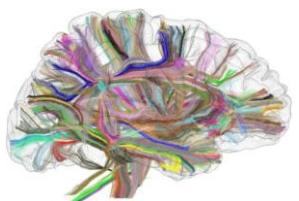
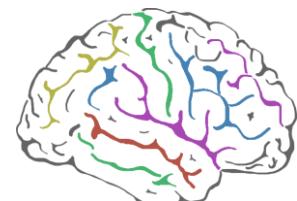


N4Brain: une infrastructure de services pour la recherche en neurosciences

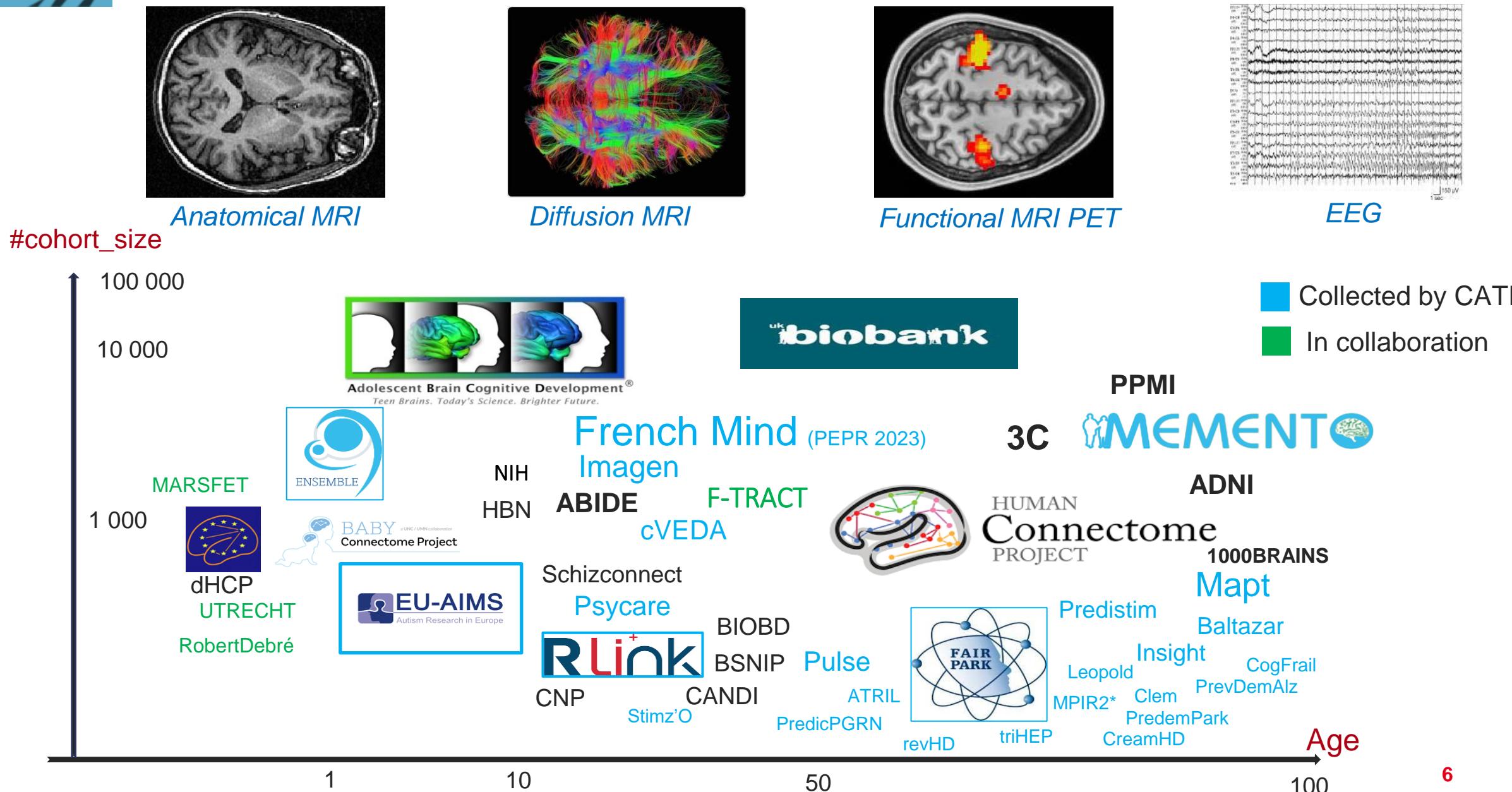


Besoins:

- Répondre aux besoins croissants des projets: PEPR PROPSY, SantéNum, MND, IHU-IRDCE...
 - EBRAINS-France (coord. CEA): fédération communauté recherche « Neurosciences » (IA/Statistique)
 - Réutilisation données recherche « grande échelle » multimodale (imagerie, génétique...), populationnelle...
 - Conformité RGPD et CNIL...
- ⇒ Capitaliser sur l'écosystème existant : infra de services CATI, logiciel Shanoir, ...
- ⇒ Mutualisation des ressources et coûts équipements/maintenance (projets Nationaux & E.U.)
- ⇒ Savoir-faire et services spécifiques aux données (pre-processing, normalisation, Quality Control, modèles IA...)
 - ❖ Neuroimagerie
 - ❖ Génétique
 - ❖ Clinique

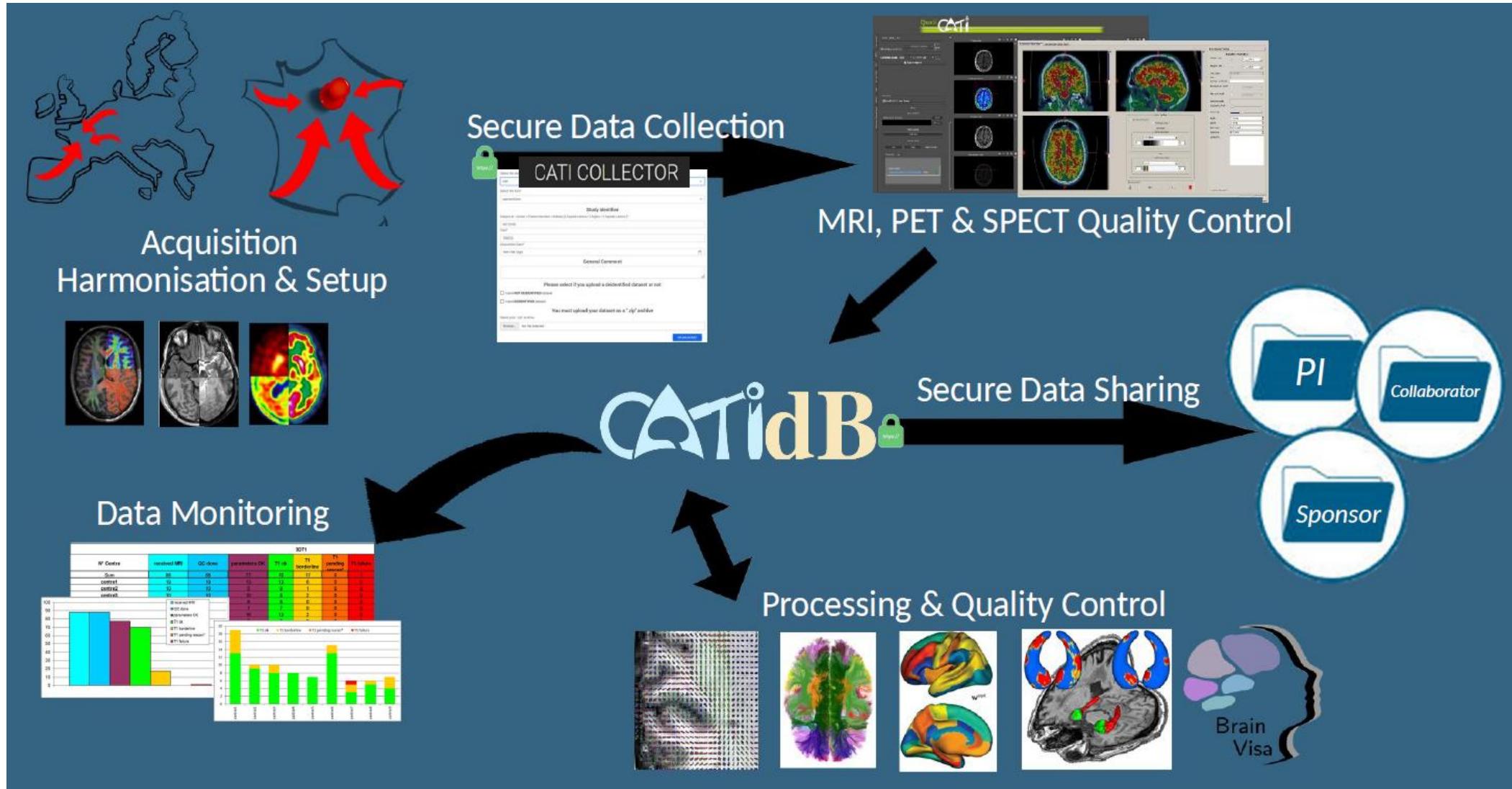
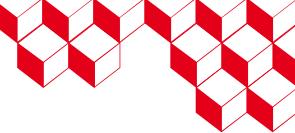


Neuroimagerie: paysage des données de l'UMS CATI (AP-HP, CEA, CNRS, INSERM, Sorbonne Université)

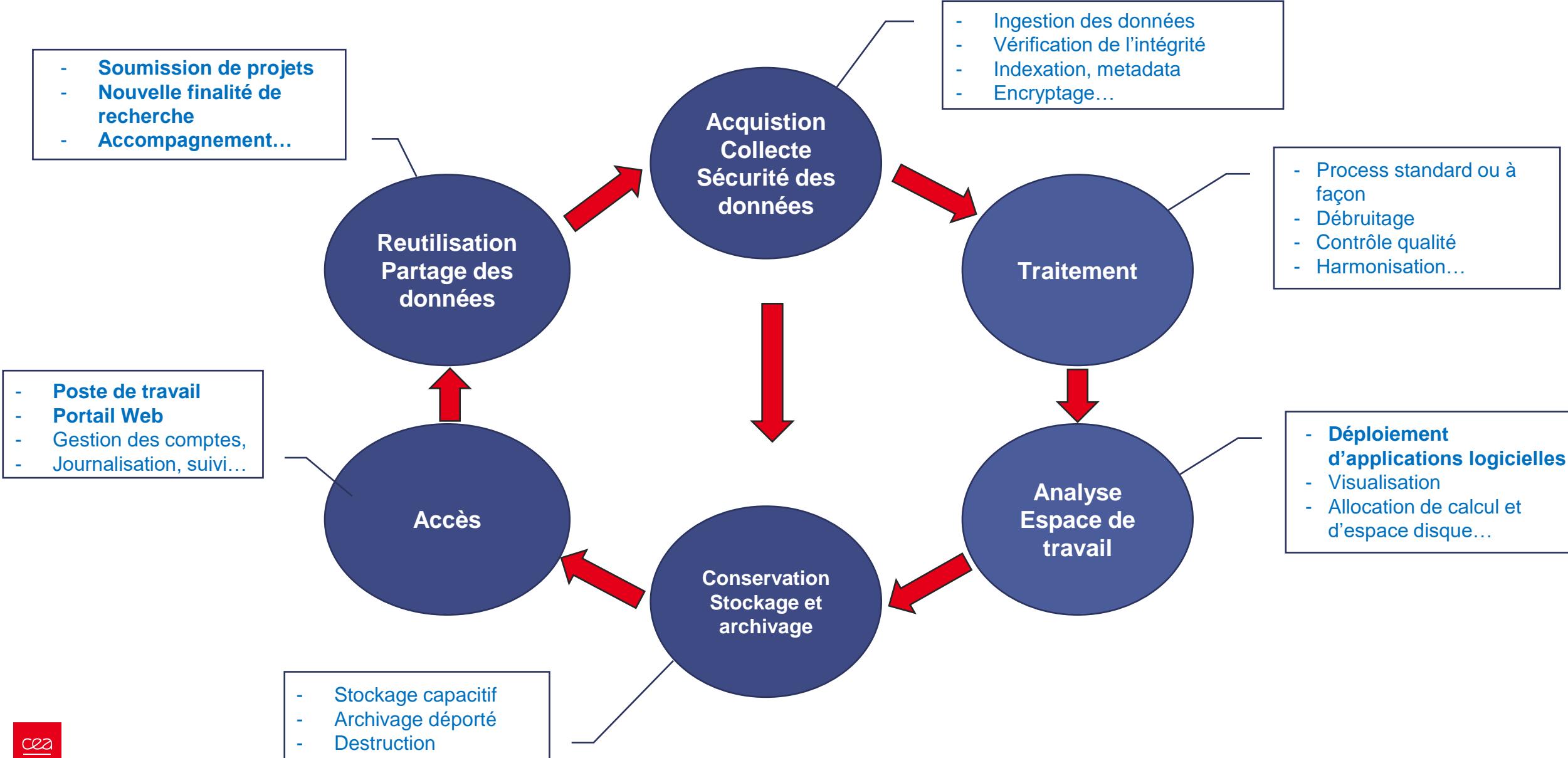


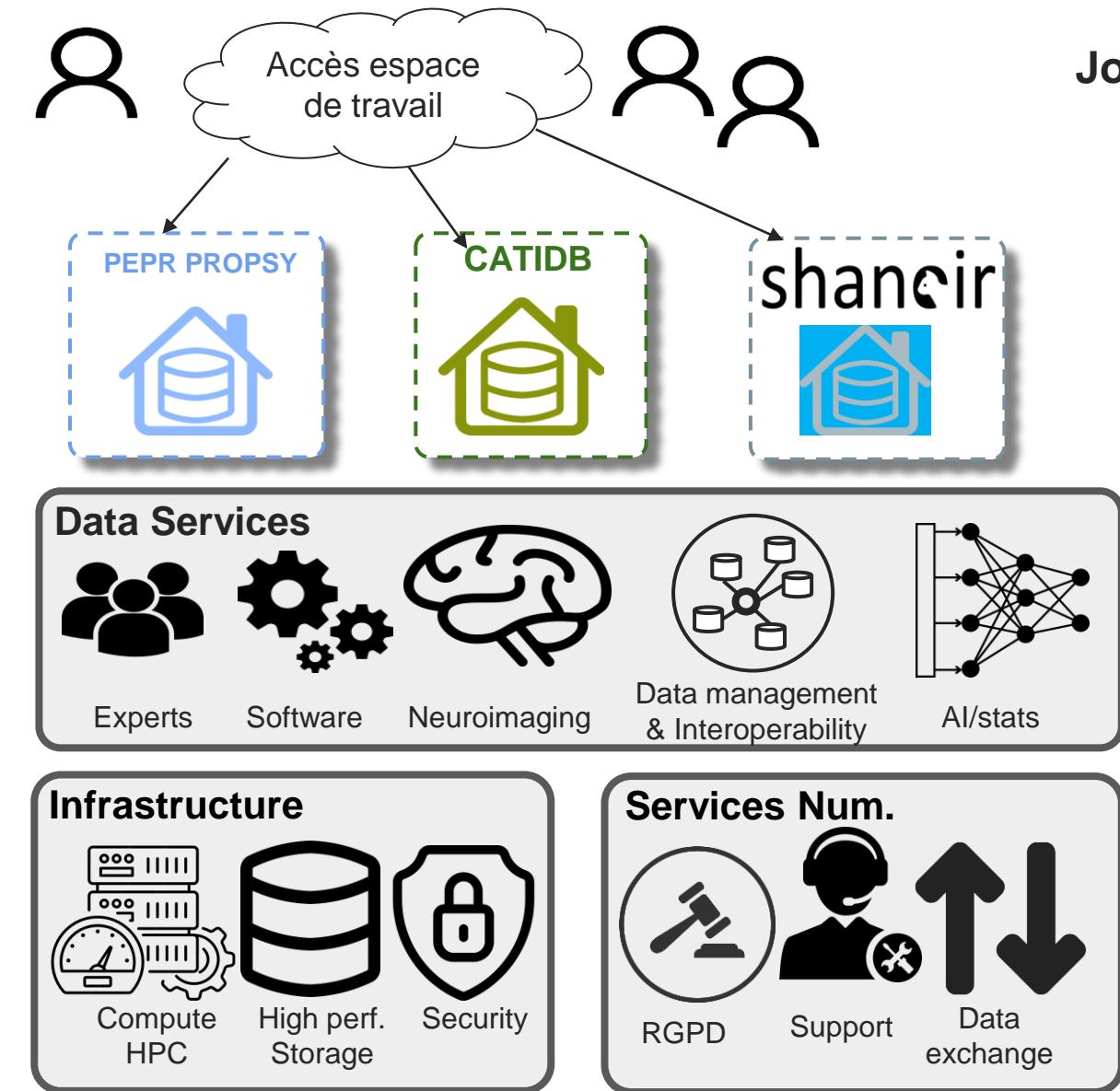
Les services aux données du CATI

=> gestion et métiers de la donnée : ~12 pers. mobilisées



Entrepôt « N4Brain » : couverture cycle de vie des données et extension registre des traitements





Joint initiative :



- Accès au catalogue de données via espaces sécurisés: CATI, PROPSY (PSYCARE, French-MIND), OFSEP, EU-RLINK, AIMS, ABCD...
- Services aux données (**Neuroimagerie et IA**)
 - Harmonisation et indexation des collections
 - Contrôle qualité et SOP
 - Gestion et interopérabilité (FAIR)
 - Analyses statistiques et IA dédiée
 - Env. logiciels « métiers »
- Services Numériques
 - Encryptage des données, cybersécurité
 - Support utilisateurs
 - Interconnexion sécurisée avec d'autres ressources (e.g. **Cloud-HDS Inserm**, PDS, CAD, EDS...)
- Infrastructure sécurisée et accès
 - Calcul haute performance (~100 GPU, 1600 CPU)
 - Stockage capacitif et backup : **2 Po => 5 Po**
 - Maintenance infra



<https://www.cnil.fr/fr/traitements-de-donnees-de-sante-comment-faire-la-distinction-entre-un-entrepot-et-une-recherche-et>

- **Exigences réglementaires (référentiel EDS CNIL) et admin:**

- ~30aine points documentaires (PSSI CEA) et/ou technique (réponses logicielles) identifiés
- Portail de transparence (gestion consentement et droit à l'information)
- Rédaction AIPD : accompagnement DPD & service de la santé CNIL
- Homologation du SI & rédaction des CGU

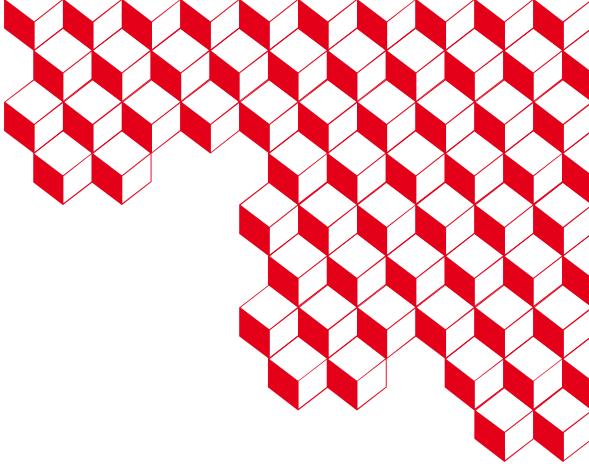


Règlement EEDS & table ronde

Constat: données de santé générées en recherche sont distinctes données produites par établissements de soin:
origine, nature et la PI associée sont **distinguantes** et **conditionnent les modalités d'accès**.

	Données Soin	Données Recherche
Volume	Massive	Moins massive
Data	Qualité variable	Hautement qualitative
Protocole	Variable	Harmonisé
Financeur	Système de santé	Institutionnels, ...

- Instruire les spécificités des données de recherche au sein du futur ORAD
- Impact sur redevance : coûts de curation (collecte, harmonisation, nettoyage, structuration, enrichissement, maintien) des données produites et valorisation métiers de la donnée ?
- Détenteur de données: vers un statut du détenteur de données de confiance ?



Merci pour votre attention