

# GRICAD : exemple de structuration et d'évolution d'un mésocentre

Violaine Louvet, Directrice GRICAD UAR3758

- En 2016 : **fusion** des universités, Idex
- **Différentes structures d'établissements** : DSIs, structure inter-universitaire
  - Et la **fermeture** d'une unité sur des périmètres très proches
- Historique de **collaborations** fortes sur des projets communs : réseau, calcul, stockage...
- Des enjeux
  - de **mutualisation** et de **rationalisation** des ressources humaines et matérielles
  - De **besoins croissants** des communautés scientifiques, avec une **diversification** des modalités numériques de plus en plus importante : calcul intensif, traitement de données, stockage, calcul à la demande, diffusion de données ...
  - Des communautés ayant une **appropriation** très diverse de l'outil numérique
  - D'**évolution technologique et matérielle** importante

- Créée au **01/01/2016**, renouvelé en 2021
- Unité d'Appui et de Recherche ayant pour tutelles : **l'UGA, le CNRS (DDOR, administrativement rattaché à l'INSMI), GINP-UGA et INRIA**



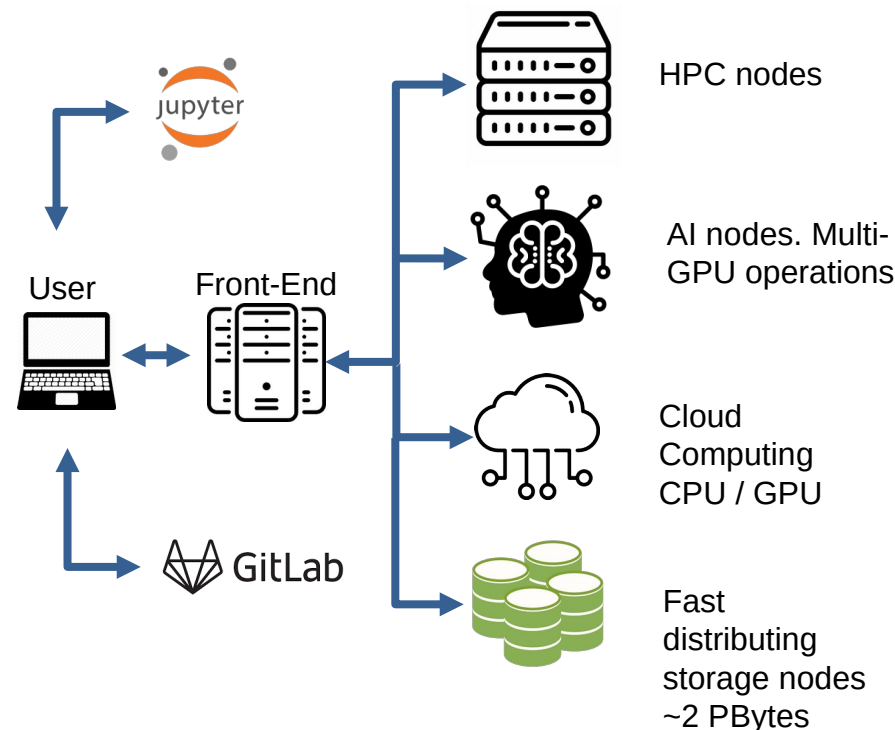
- **25 personnes** (au 21/10/2022) dont
  - 17 permanents affectés GRICAD
  - 4 postes permanents mutualisés avec des labos (un 50/50, deux 20/80, un 25/75)
  - 4 CDD (3 PNRIA et un remplacement départ mobilité)

## Modèle organisationnel mixte

- Des équipes en **interaction forte** : laboratoires, DSI
- **Mutualisation** des moyens humains sur les projets d'infrastructures de site
- Favorise la **proximité** avec les équipes de recherche, facteur de montée en **compétence**

- **Comité de pilotage** composé des représentants des tutelles de l'unité
  - Définition des orientations de l'UAR
  - Budget et demande de moyens
- **Comité de direction** composé de la direction, adjoint à la direction, responsables d'équipes, responsable financière, responsable RH et communication
  - Fonctionnement quotidien opérationnel
  - Mise en oeuvre des orientations définies par le comité de pilotage
- **Assemblée Générale** de l'UAR
  - Avis sur le fonctionnement de l'UAR, l'organisation ... toute question relative à la vie de l'unité
- **Comité Scientifique et Techniques des Utilisateurs (COSTU)** composé d'utilisateurs des services de l'UAR représentatifs des communautés et des usages
  - Expression des besoins des communautés
  - Retours sur les usages, difficultés ...
  - Aide au pilotage scientifique des services

- **Accompagnement et conseils** aux chercheurs sur tous leurs besoins liés aux données et au calcul
- Mise à disposition de l'ensemble des chercheurs et personnels en support de la recherche d'**infrastructures avancées et mutualisées** pour l'exploitation des données de la recherche et le calcul intensif.



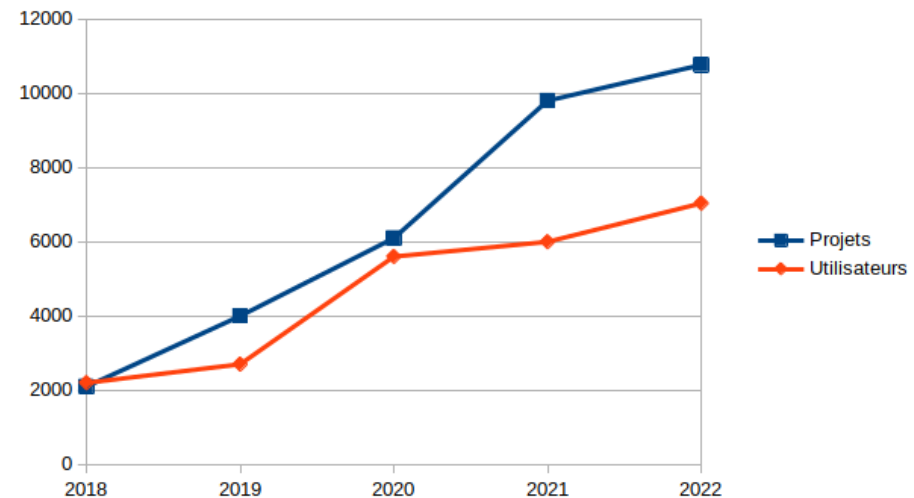
- Structurellement : **postes partagés** avec plusieurs laboratoires
- Des comités techniques **multi-partenaires** qui assurent l'exploitation des infrastructures :
  - Stockage
  - Hébergement serveurs / DC grenoblois
  - Clouds
  - HPC
- **Multi-partenaires** = GRICAD, DSI UGA, laboratoires
  - Responsabilité GRICAD et financement recherche sur les infras périmètre recherche (HPC, cloud, ...)
  - Responsabilité DSI et financement établissement sur les infras plus génériques (stockage, hébergement, cloud, ...)



\* Services DGDSI/UGA

- **HPC / Cloud** : 218 projets actifs, 846 utilisateurs actifs
  - Dont 70 projets cloud (143 instances, 2.3 TB utilisés sur 4.5 TB, 767 VCPU utilisés sur 1056)
  - 34 millions d'heures de calcul pour l'année écoulée (+8 millions par rapport à l'année précédente)
- **Répartition par pôle** :
  - MSTIC (Mathématiques, sciences et technologies de l'information et de la communication) : 28,44 %
  - PAGE (Physique des particules, astrophysique, géosciences, environnement et écologie) : 36,70 %
  - PEM (Physique, ingénierie, matériaux) : 18,80 %
  - CBS (Chimie biologie santé) : 14,22 %
  - SHS (Sciences humaines et sociales) : 0,46 %
  - PSS (Sciences sociales) : 1,38 %
- **Stockage performant** : 1,08 Pb utilisé sur 1,47 Pb

- **Stockage semi-pérenne** : 191 Tb utilisé sur 2 Pb (infrastructure récente)
- **Jupyter** : + de 3500 utilisateurs depuis le lancement.
  - 800 Go de stockage
  - 50 à 100 utilisateurs par jour dont 30 à 60 en simultané



Evolution de l'utilisation de la plateforme gitlab



- Dynamique d'évolution **intrinsèque à nos activités** :
  - Évolution des **besoins utilisateurs**
  - Évolution de l'**écosystème et des contraintes**
  - Évolutions **technologiques**
- Indispensable d'**accompagner ces changements**
  - Importance de la **veille technologique**, de la mise en place de prototypes et de nouvelles infrastructures en capacité de répondre à des nouveaux besoins
  - Intégrer les dynamiques autour des **données de recherche, de la science ouverte**
- ➔ Le numérique pour la recherche est un **continuum multi modal** : il est important de considérer les problématiques dans leur globalité.

Depuis quelques années, au quotidien pour GRICAD et les personnels d'appui à la recherche dans les laboratoires :

- Des interrogations autour du stockage
- Des questions autour du partage des données, de leur sécurisation, du RGPD, ...
- Des demandes pour la diffusion des données, la rédaction de Plan de Gestion de Données, ...

De plus en plus **nombreuses**.

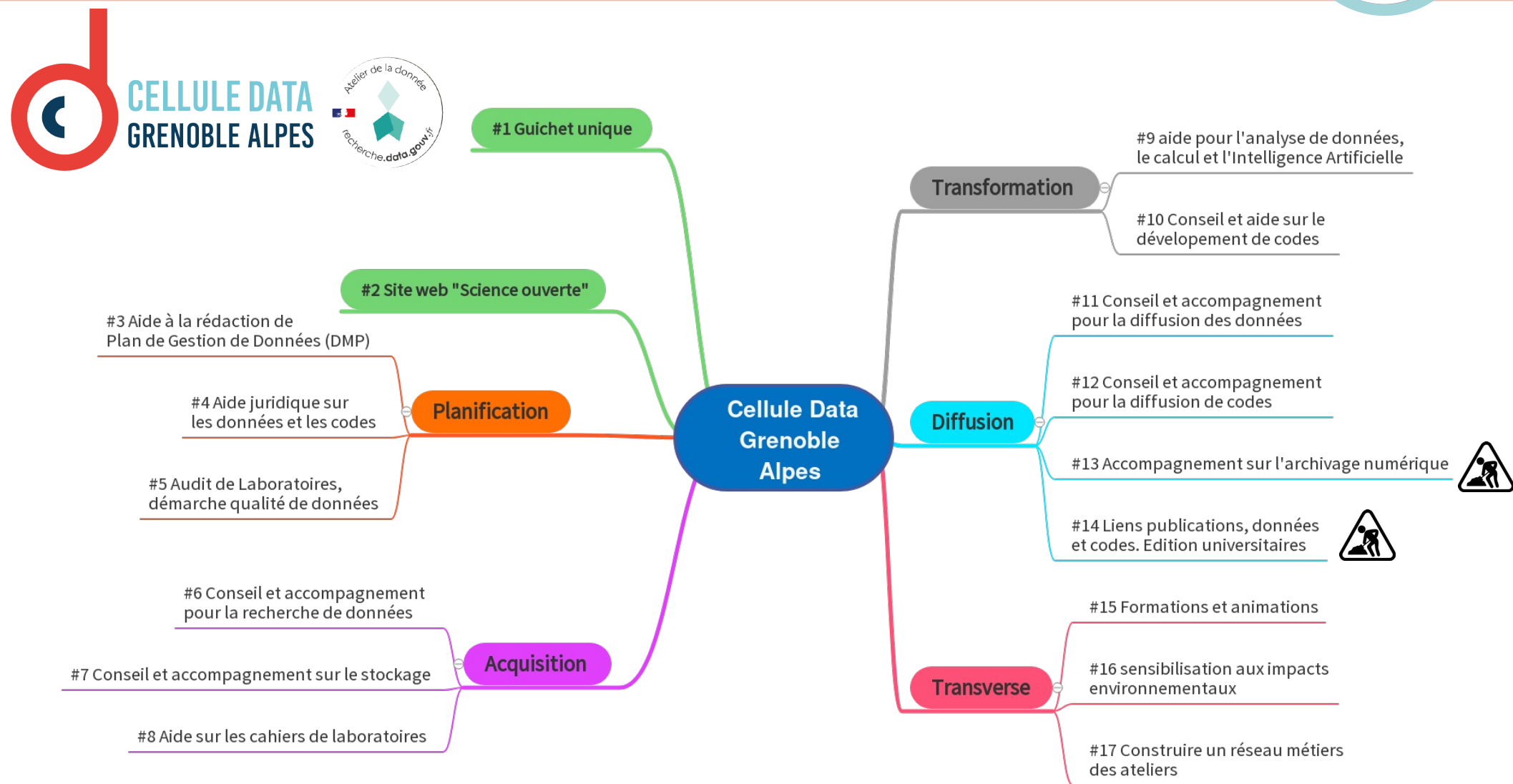
En parallèle, sollicitation des **personnels des bibliothèques** sur la diffusion, la description et le signalement des données.

## Comment répondre efficacement et globalement à toutes les problématiques ?

- Nécessité de **rassembler des compétences complémentaires**
- Dans un environnement national, européen et international extrêmement **dynamique**.
- Avec des **obligations** qui évoluent pour les scientifiques : Plan de Gestion de Données, données ouvertes ...

- Mise en place d'une **cellule data multi-partenaires opérationnelle**, coordonnée par GRICAD composée de 23 membres :
  - GRICAD (6)
  - Bibliothèque - Appui à la Science Ouverte (BAPSO) (7)
  - Laboratoires (dont GRESEC et Litt&Arts, labos SHS) (6)
  - SED INRIA (3)
  - Délégué à la Protection des Données de l'UGA
  - MSH Alpes (Ingénieur PUDGA en cours de recrutement)
- Des personnels à la fois **UGA , CNRS, CEA et INRIA**
- Des **compétences complémentaires** : informatiques, documentation, juridiques, science ouverte, scientifiques
- Depuis 2019, propose un ensemble de **services autour de la donnée** en s'appuyant sur les infrastructures du site
- S'appuie sur un **réseau de référents données** dans les laboratoires





- Un **site web Science ouverte Université Grenoble Alpes**
  - › Rubriques publications, données, codes et logiciels
  - › Services et supports de formations
  - › Actualités et agenda
- Des actions de **formation et d'animation** auprès de différents publics (chercheurs, personnels techniques, doctorants...) sur les thématiques liées aux données de la recherche :
  - › **Formation des doctorants**
  - › **Webinaires**
- Informer et accompagner sur les **outils, services et infrastructures locaux ou nationaux** (par exemple HumaNum, Progedo, GENCI) répondant de la façon la plus pertinente aux besoins décrits
  - › Conseils autour des architectures techniques, des logiciels, des bases de données
  - › Accompagnement au stockage, sécurisation, chiffrement, calcul, IA ...
- Aide à la rédaction de **plan de gestion des données, analyse d'impact**
- Accompagnement des **politiques de laboratoire** : demande de quelques laboratoires pour les aider dans leur réflexion sur la mise en place d'une politique globale des données

- Participation importante (avec l'implication de 4 personnes du site) à l'**entrepôt national recherche.data.gouv**
  - **Recherche Data Gouv : une solution souveraine** pour permettre aux chercheurs de déposer leur données de recherche lorsqu'ils n'ont pas d'entrepôt de confiance dans leur discipline.
  - Organisé en **espaces institutionnels** de publication et de signalement des données des établissements qui souhaitent participer et d'un espace générique pour les chercheurs dont l'établissement n'a pas de collection
    - ➔ L'UGA a fait le choix de disposer d'une collection établissement sur Recherche Data Gouv.
  - **Modération des dépôts dans la collection UGA** par la CDGA et des membres de laboratoire

# Un écosystème au service du partage et de l'ouverture des données de recherche

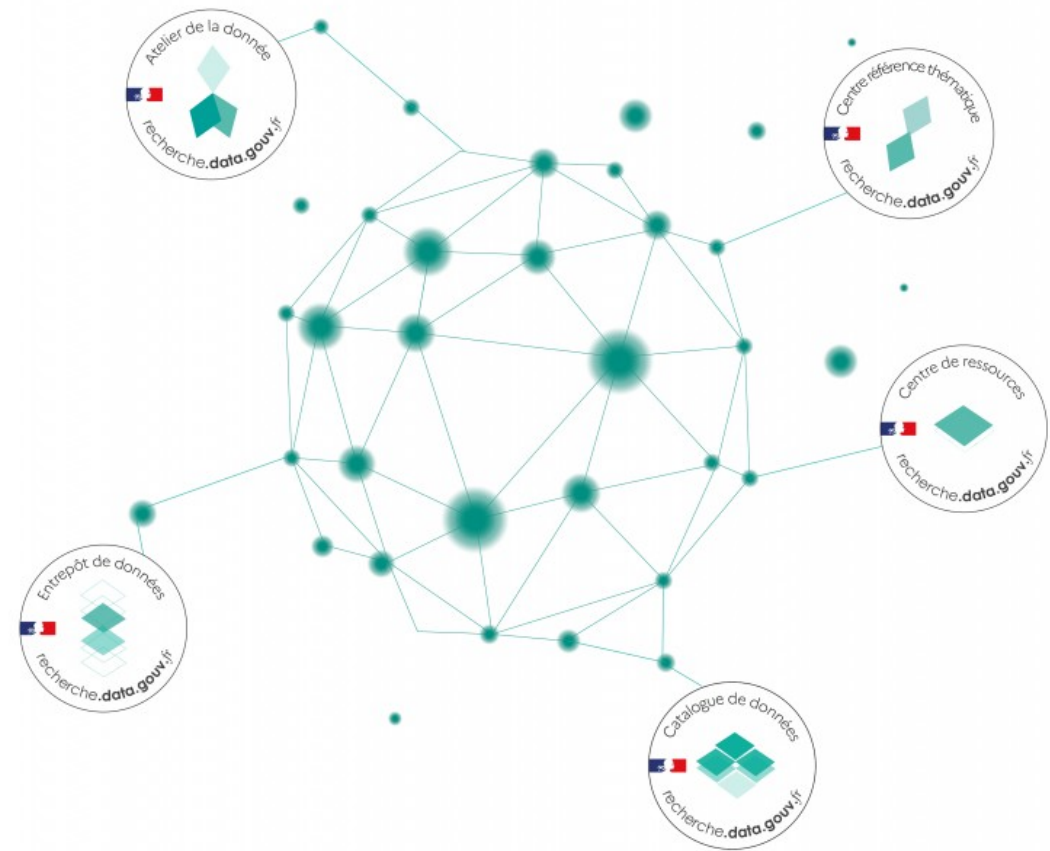
FÉDÉRER, ACCOMPAGNER, PARTAGER, OUVRIR, RÉUTILISER

## Trois modules pour accompagner les équipes de recherche sur toute question relative aux données :

- Des ateliers de la donnée
- Des centres de référence thématiques
- Des centres de ressources

## Deux modules pour déposer, publier et signaler des données :

- Un entrepôt pour déposer et utiliser des données
- Un catalogue pour rechercher les données publiées sur l'entrepôt ou sur des entrepôts externes



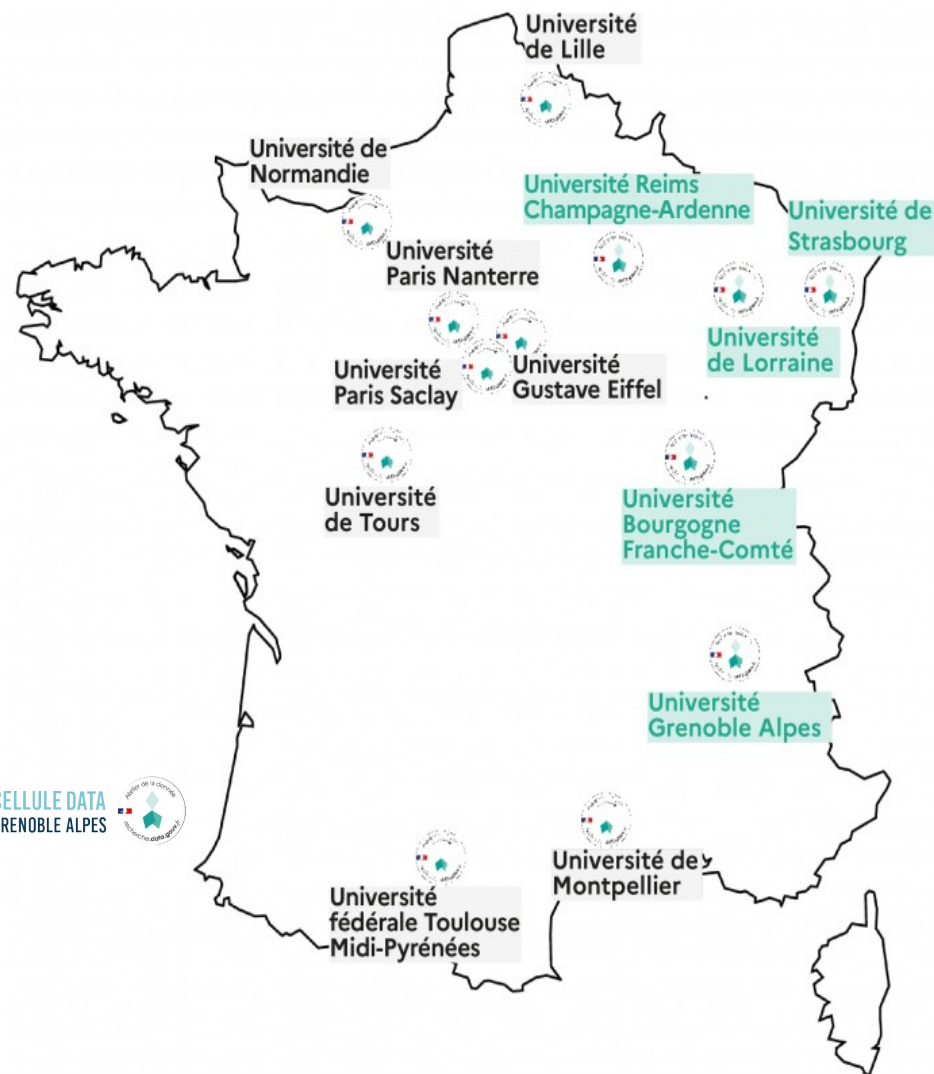


# Un écosystème au service du partage et de l'ouverture des données de recherche

FÉDÉRER, ACCOMPAGNER, PARTAGER, OUVRIR, RÉUTILISER

## Volet atelier de la donnée

- **Objectif** : labelliser les initiatives existantes ou nouvelles sur l'accompagnement des communautés scientifiques tout au long du cycle de vie des données.
- En proximité, regroupant des compétences complémentaires
- Différents niveaux de labellisation selon la maturité et le niveau de service apporté
- 5 ateliers labellisés dont la **Cellule Data Grenoble Alpes** !!
- Constitution en cours d'un **réseau des ateliers** pour se coordonner et faciliter les échanges, basé sur le réseau initial constitué avec les collègues de l'Est (Lorraine, Strasbourg, BFC, Champagne Ardennes et Grenoble)





- Poursuivre le **travail de terrain**
  - Accompagner, communiquer, informer
  - Soutenir la dynamique d'évolution et la veille technologique
  - Dans les laboratoires, dans les équipes mais aussi auprès de la gouvernance
- Intégrer les ressources du **projet Excellences dans la Cellule Data Grenoble Alpes**
  - Mise en place d'un pool d'ingénieurs pour l'accompagnement des laboratoires SHS sur toutes les problématiques des données de recherche
- Accompagner les **projets structurants nationaux** (MesoNet, DIADEM, Gaia Data) et **locaux** (Idex : Cross Disciplinary Project et CDTools)
  - Pousser les outils de reproductibilité comme Guix et Nix au niveau des mésocentres
- Assurer la validation du **modèle économique** auprès des tutelles
- Poursuivre les travaux autour de l'**impact environnemental**
  - Discussions en cours sur les choix en cas de pic de consommation mais aussi de façon plus générale