



# La Cellule Data Grenoble Alpes

JCAD 2021, 15 décembre 2021

# Quelques rappels sur GRICAD



- Créé au **01/01/2016**, renouvelé en 2021
- **Unité d'Appui à la Recherche** ayant pour tutelles : **UGA (dont G-INP)**, **CNRS (MICADO/DDOR, administrativement rattachée à l'INSMI)** et **INRIA**



## → Missions :

- **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.



Au quotidien pour GRICAD et les personnels d'appui à la recherche dans les laboratoires, mais aussi les personnels des bibliothèques :

- Des interrogations autour du stockage
- Des questions autour du partage des données
- Des demandes pour la diffusion des données ...

De plus en plus **nombreuses**.

- Dans un environnement national, européen et international extrêmement **dynamique**.
- Avec des **obligations** qui évoluent pour les scientifiques : Plan de Gestion de Données, Science ouverte, données ouvertes ...
- Et un volume de données à gérer en constante augmentation



- Disponibilité de **très gros volumes de données** issues de sources **très variées**
- Les **méthodes de travail**, tout comme les **technologies d'analyse des données**, ont fortement **évolué** : travail collaboratif et nécessité de partage, émergence de l'open data, outils d'indexation et de recherche, explosion de l'IA ...
- **Diversification des besoins** : traitement de données, intelligence artificielle, calcul intensif, stockage, diffusion et valorisation de données, questions juridiques autour des données, architecture technique et base de données ...
- **Accompagnement à adapter**, communautés “neuves” dans le domaine, connaissance et appropriation technique très hétérogènes des technologies et des infrastructures d'une communauté à l'autre, d'un individu à l'autre
- De plus en plus d'**interdisciplinarité** : besoin de partage et d'organisation de données de sources diverses pour des communautés très différentes
- Un soutien technique dans les laboratoires **très inégal** d'une structure à l'autre

# Un cadre qui évolue fortement



- **RGPD** (2016) : règlement européen sur la protection des données personnelles
- **Obligation de Plan de Gestion de Données** exigée par de plus en plus de financeurs de la recherche (H2020, ANR, Horizon Europe, ...)
- Loi pour une république numérique (2016), Plan National pour la Science Ouverte (2018, 2021), H2020 (depuis 2017) et Horizon Europe (à partir de 2021), Feuille de route de la science ouverte du CNRS (2018), Plan données de la recherche CNRS (2020) : **l'ouverture des données de recherche financées sur fonds publics devient la norme**



# Les enjeux d'une bonne gestion des données de la recherche



- En premier lieu **faciliter le travail de recherche** lui-même, mais les enjeux sont plus larges :
  - Assurer la **conservation** des données à moyen terme
  - Pouvoir **reproduire** les résultats scientifiques, y compris par l'équipe qui les a obtenus !
  - Pouvoir facilement les **réutiliser** pour produire de nouvelles recherches
  - **Valoriser** les résultats scientifiques, mais aussi les données et les codes qui ont permis de les obtenir et en augmenter la visibilité
  - Favoriser de **nouvelles collaborations**, de nouvelles approches
- Enjeux d'ordre sociétaux :
  - Assurer une **souveraineté** sur les données produites
  - Assurer l'**intégrité scientifique**
  - Garantir la **transparence**, et assurer la confiance des citoyens en la recherche



“L’Europe me demande un DMP, qu’est-ce que c’est ?”

“Je n’arrive plus à ouvrir mon tableur avec ce fichier de données”

“Je veux utiliser AWS, comment je fais ?”

“Je stocke mes données sur un DD externe qui se trouve dans mon bureau”

“Je veux diffuser mes données sur le web”

“J’ai pas mal de données de types différents et je pense que ce serait très intéressant de pouvoir les croiser”

“Comment diffuser mon code avec mes données ?”

“On doit partager des données entre plusieurs collègues dont données soient diffusées”

“Le financeur de mon projet demande à ce que mes données soient diffusées”

“J’ai loué de la volumétrie pour mon projet mais celui-ci est terminé et je n’ai plus d’argent”

“Mon équipe aimerait tester le Deep Learning sur nos données”



- Des structures en soutien avec des expertises différentes :
  - GRICAD
  - BAPSO
  - MSH
- Pas mal de laboratoires concernés et très impliqués

## La Cellule Data Grenoble Alpes - CDGA

Des compétences complémentaires : **techniques, science ouverte, juridique.**

16 membres issus de :

- GRICAD
- BAPSO, DDOR (CNRS)
- MSH
- Labos
- + le DPD du site





- Une structure opérationnelle pour :
  - Répondre **concrètement** à tous les questionnements des scientifiques
  - Fournir un **point d'entrée unique** aux communautés
  - Constituer et animer un réseau de **référents** autour des données pour chaque laboratoire
  - Mettre en place les **outils, services et infrastructures** répondant aux besoins des communautés
  - **Animer, former** (chercheurs, personnels techniques, doctorants...) sur les thématiques liées aux données de la recherche
  - Faire de la **veille juridique et technique**, et s'inscrire dans les initiatives nationales, européennes et internationales



## **EOSC, European Open Science Cloud**

- Assurer l'**interopérabilité et le partage** des données au niveau européen
- En s'appuyant sur les infrastructures et services existants

## **RDA, Research Data Alliance**

- Démarche **bottom up**.
- Accélérer et faciliter le partage et l'échange des données scientifiques
- De nombreux groupes de travail et d'intérêts
- Des recommandations et produits de référence
- Un **noeud RDA France**

Assurer la présence grenobloise dans ces projets, faire le lien avec les communautés scientifiques du site.



- Les **coûts humains** liés aux données de recherche sont très importants, très sous-estimés et très peu pris en compte
- Ils ne pourront pas reposer sur un petit ensemble de personnes mais devront se répartir sur l'**ensemble du collectif**

Outre l'accompagnement pratique, l'un des objectifs de la cellule data est d'accompagner les changements **culturels, méthodologiques, professionnels**, liés aux données :

- de proposer des **actions de sensibilisation** sur ces questions
- d'organiser des **formations** au niveau du collège doctoral et à destination des chercheurs et ingénieurs
- de proposer des **séminaires, ateliers, retours d'expérience ...**

# Quelques exemples : autour des Plans de Gestion de Données



- Réponses à toutes les demandes concernant les **plans de gestion des données** et les documents nécessaires dans le cas de données sensibles (RGPD / AIPD).
- Relecture, aide à la rédaction.
- Rédaction de documents pour les aspects techniques liés à l'utilisation des infrastructures du site.
- Aide sur la documentation des données et sur les métadonnées (identification des standards de métadonnées, ...).
- Interface avec les services de valorisation.
- Aide à l'utilisation de DMP Opidor.
- Aide à l'élaboration de la stratégie de valorisation et le cas échéant identification du ou des entrepôts les plus adaptés.
- Mise en place de formations et d'ateliers. Rédaction de fiches pratiques.

# Quelques exemples : accompagnement sur le stockage



- Réponses à toutes les demandes concernant la **problématique stockage** :
  - Orientation vers la plateforme adéquate.
  - Aide au partage des données entre partenaires d'un projet.
  - Aide à l'organisation des données.
  - Aide à la sécurisation des données.
  - Accueil et encadrement de CDD data scientists des projets.
  - Rédaction de documentations et de recommandations.
  - Mise en œuvre d'architectures techniques pour certains projets.
  - Mise en place de formations et d'ateliers. Rédaction de fiches pratiques.

# Quelques exemples : formations et animation



- **Formations** organisées à différents niveaux :
  - › collège doctoral,
  - › équipes de recherche, personnels d'appui à la recherche et chercheurs.
- Sur de **nombreux sujets** :
  - › Plans de Gestion des Données
  - › génie logiciels,
  - › calcul intensif,
  - › IA,
  - › stockage,
  - › diffusion des données, ...
- Interventions dans des formations **disciplinaires**, et au niveau **national**.
- Organisation de **séminaires et d'ateliers**.
- Animation d'un **réseau de référents dans les laboratoires**.
- Animation d'une **liste de diffusion**.
- Création de **ressources** (fiches, vidéos ...) sur différents sujets.



- Participation importante au projet **Recherche.Data.Gouv** dans son module « Entrepôt »
  - Collection propre UGA
- Réponses aux demandes pour la **diffusion en open data** des données de recherche.
- Orientation vers les **entrepôts les plus adaptés**.
- Conseils sur les **formats**, l'**organisation**, la **documentation** des données, les standards de **métadonnées**.
- Aide pour la rédaction de **data papers**.



- Pour contacter les membres de la cellule
  - [uga-cellule-data@univ-grenoble-alpes.fr](mailto:uga-cellule-data@univ-grenoble-alpes.fr)
- Pour être aider :
  - [sos-data@univ-grenoble-alpes.fr](mailto:sos-data@univ-grenoble-alpes.fr)
- Pour se tenir informer : abonnement sur la liste uga-research-data
  - <https://listes.univ-grenoble-alpes.fr/sympa/info/uga-research-data>
- Site web :
  - <https://scienceouverte.univ-grenoble-alpes.fr/>

