



Séminaire :
**Accompagner les doctorants sur la
gestion des données**

Mardi 05 Juillet 2022



- 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
 - Assurer la présence grenobloise dans les **projets nationaux et européens**, faire le lien avec les communautés scientifiques du site

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

21 membres (UGA, CNRS, CEA, INRIA) issus de :

- GRICAD
- BAPSO
- Labos (dont CEA)
- SED INRIA
- + la DPD (Déléguée à la Protection des Données) du site



Quelques exemples d'**accompagnement tout au long du cycle de vie des données et des codes**

- Mise à disposition d'**outils de travail collaboratifs** sécurisés
- Aide à **trouver** des données et des codes
- Aide au **stockage / au traitement** des données
- Aide à la **description / la diffusion / l'archivage** des données et des codes
- **Lien publications/données/codes**
- Service de carnets de laboratoires numériques (**ELN**)
- Aide **juridique**
- Aide à la rédaction de **plan de gestion des données**
- **Formations**
- Diffusion de **bonnes pratiques**



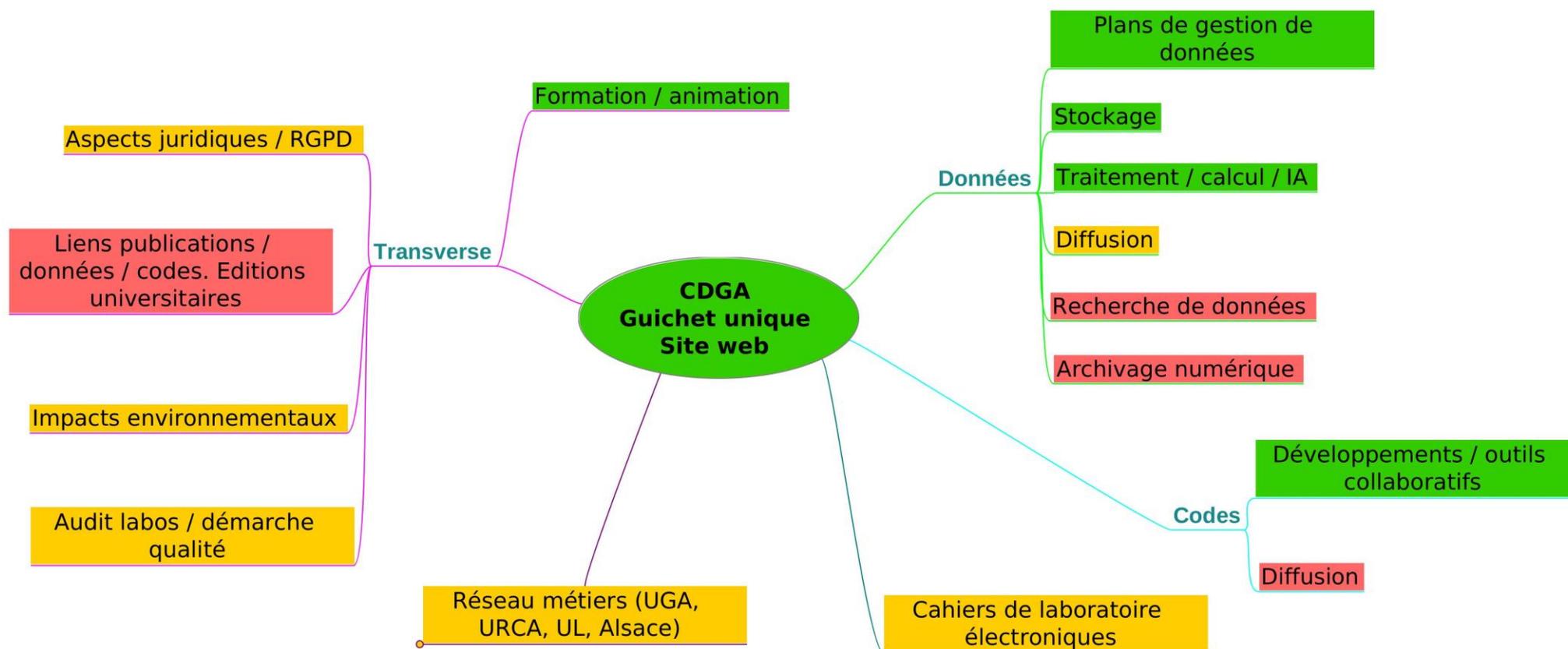
- **Pour résumer :**
- Un **point d'entrée unique** :
 - Site web : <https://scienceouverte.univ-grenoble-alpes.fr/>
 - Support : sos-data@univ-grenoble-alpes.fr
- Des **formations et séminaires**
 - [Formation CED](#) (2022)
 - Séminaires : stockage, PGD
- La participation active au projet [Recherche Data Gouv](#) et une **collection UGA** hébergée sur cette plateforme
- La mise en place d'un [réseau de référents données](#) dans les laboratoires
- Un **accompagnement au quotidien** sur toutes les questions autour des données : relecture / aide à la rédaction de PGD, accompagnement sur le stockage, sur les architectures techniques, questions juridiques, ...

La Cellule Data Grenoble Alpes



La cellule data Grenoble Alpes [labellisée](#)
« [atelier de la donnée](#) » au niveau national

- Vert : Service en place
- Orange : Service existant en cours de consolidation
- Rouge : Nouveau service à initier





Si votre laboratoire n'a pas de référent données : souhaiteriez vous être référent données ?

La liste des laboratoires ayant un référent donnée est présente au bas de la [page web de la Cellule Data](#).



**Enquête sur les usages et les besoins pour la gestion
des données de la recherche sur le site de l'Université
Grenoble Alpes (2020-2021)**

Focus doctorants



- Convergence de grands enjeux liés :
 - à l'**augmentation exponentielle** des données dans tous les domaines scientifiques
 - à l'amplification du mouvement d'**ouverture de la science**
- Modifications profondes des **processus** de recherche et des **méthodologies** de travail des communautés scientifiques
- **Bonne gestion des données de la recherche** : une nécessité pour faire face à ces défis :
 - facilite le travail de recherche, individuel et collectif
 - sécurise les fruits de la recherche
 - simplifie leur diffusion

Objectifs : Réaliser un **état des lieux des pratiques** autour des données de la recherche auprès de toutes les communautés scientifiques du site de Grenoble de façon à **adapter l'accompagnement** dans cette dynamique de changement extrêmement rapide



- Questionnaire proposé de **novembre 2020 à janvier 2021**
- Diffusion très large sur le site
- **877 réponses au total** dont 421 réponses complètes et 456 réponses partielles. Au total **578 réponses** complètement exploitables et des éléments supplémentaires des autres réponses
- Réalisé en collaboration avec des étudiants de M2 MDLE information et documentation
- Les résultats complets de l'enquête sont disponibles sur [le site Science Ouverte](#) de l'UGA



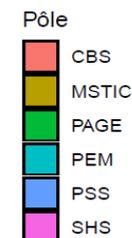
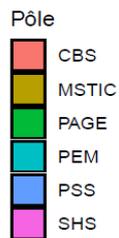
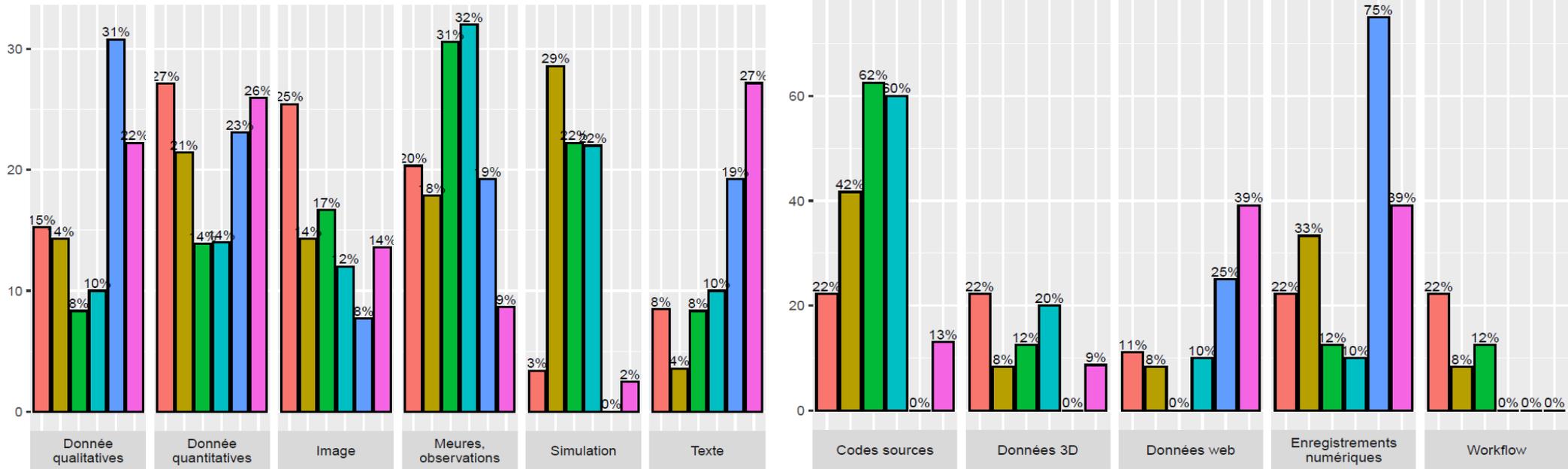
Profils des répondants

- Les pôles CBS, MSTIC, PAGE et PEM (entre 13 et 19% des répondants doctorants) sont à peu près représentés de façon équivalente dans les répondants doctorants. Le pôle SHS est plutôt sur-représenté (30%) alors que le pôle PSS l'est beaucoup moins (7%)
- Cette répartition a un impact sur certaines des analyses proposées



Type de donnés

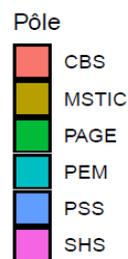
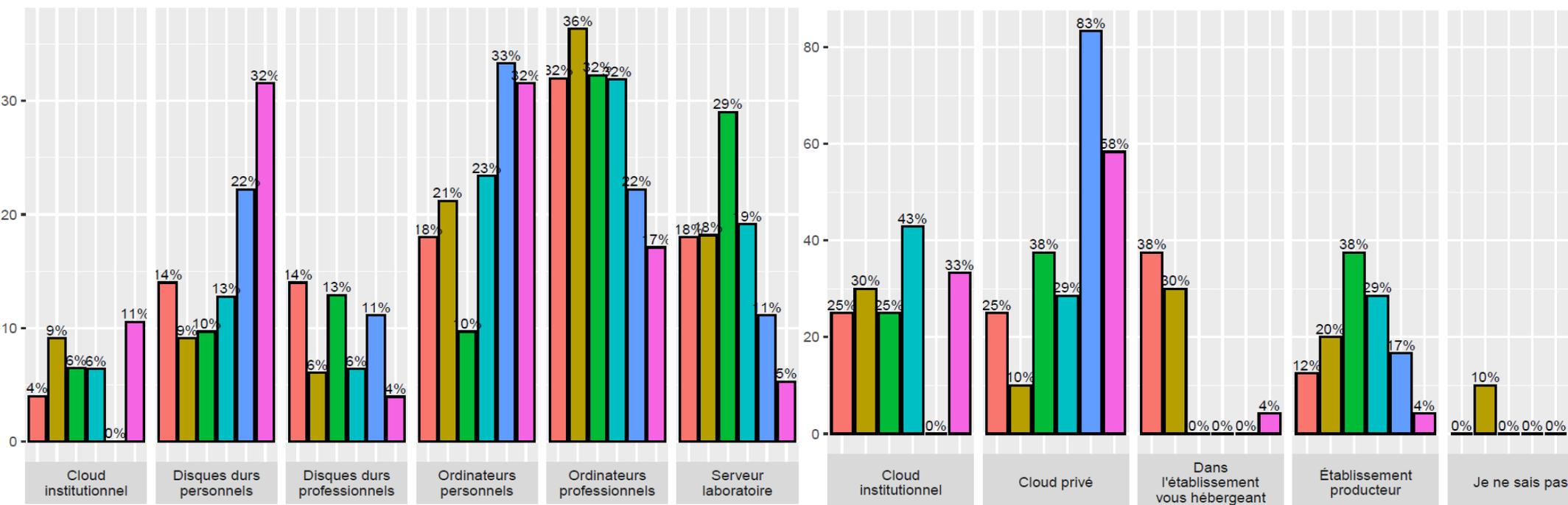
- **Objectifs** : comprendre les données qui sont manipulées par les doctorants, ainsi que des pratiques en matière de stockage



Type de données - Grandes tendances :

- les simulations et codes sources sont utilisées plutôt parmi les doctorants des pôles MSTIC, PAGE, PEM
- les textes, les données web, les données de réseaux sociaux et les enregistrements numériques audio et vidéo sont utilisés plutôt parmi les doctorants du pôle SHS
- les données qualitatives sont souvent mentionnées par les doctorants du pôle PSS
- les données de mesures ou d'observations sont citées plus spécifiquement par les doctorants des pôles PAGE et PEM

Problématique du stockage



L'utilisation d'ordinateurs personnels/professionnels et des disques durs externes personnels/professionnels sont les pratiques plus courantes

Grandes tendances :

- Les services de cloud privés (Dropbox, Google Drive ...) sont utilisés de façon non négligeable dans les pôles PSS et SHS.
- Les serveurs dédiés du laboratoire sont utilisés surtout dans les pôles PAGE et PEM.
- L'usage de matériels personnels est assez important, en particulier dans les pôles SHS, PSS et dans une moindre mesure PEM



Pratiques durant la recherche

- Environ 65% des doctorants ont déclaré avoir déjà réutilisé des données publiées par d'autres chercheurs. Il y a une répartition entre tous les pôles avec une légère majorité dans les pôles PEM et SHS puis CBS, de l'ordre de 20%, les autres pôles étant à environ 10%.
- Parmi les doctorants qui ont réutilisé des données, la très **grande majorité (88%)** a affirmé avoir eu des problèmes, en particulier :
 - données inexistantes ou impossible à récupérer sur demande
 - problème d'accès aux données
 - problème pour retrouver des donnéeset ce quel que soit leur pôle de rattachement



Personnes ressources

- Environ 44% des doctorants ont répondu ne pas connaître de personne ressource pour les aider. Les autres sollicitent quelqu'un de l'équipe ou du laboratoire (27% et 24% respectivement)
- Ces réponses sont assez similaires quelque soit le pôle d'appartenance
- Ce qui ressort des réponses ouvertes est que les doctorants demandent surtout de **l'aide à leurs directeurs de thèse et aux collègues de leur propre équipe ou du laboratoire**



Diffusion des données

- La pratique de diffuser ses propres jeux de données n'est pas très courante (14%). Pour ceux qui ont déjà diffusé des jeux de données, les lieux de diffusion cités sont plutôt **des articles**, ou des plateformes comme **gitlab**

Archivage

- Les doctorants déclarent avoir besoin d'aide pour archiver les données de façon pérenne
- Pour la pratique de l'archivage pérenne des données, les réponses libres sont assez disparates, ce qui montre un besoin de clarification des notions



Besoins exprimés

- La priorité est la formation en terme de **gestion, de publication et de citation** de données. Suivent l'aspect stockage et archivage de données et les problématiques juridiques
- Les solutions numériques dont les doctorants ont besoin pour la gestion de données en priorité sont celles associées au **stockage, à la sauvegarde et à l'archivage pérenne**
- Concernant les caractéristiques pour un **entrepôt**, les priorités sont surtout :
 - la définition par sujet et par thème
 - la certification
 - l'attribution de DOI
 - l'exploitation des données



Synthèse

- Même si les grandes tendances sont celles de l'ensemble des réponses au questionnaire, il ressort des spécificités liées :
 - au **manque d'information** sur les différentes ressources auxquels les doctorants peuvent avoir accès, tant sur les possibilités d'accompagnement que sur les plateformes et services disponibles
 - au fort **besoin de formation** sur toutes les questions liées aux données
- Il est important de noter que les services existants sur le site sont d'abord à destination des membres permanents des laboratoires, et difficilement utilisables par des doctorants



**Formation doctorale sur la gestion
des données de la recherche (mai 2022)**

Thèmes abordés lors de la formation doctorale sur la gestion des données



Objectifs :

- Sensibiliser à la gestion des données de la recherche (principes FAIR).
- Sensibilisation aux enjeux de l'ouverture des données
- Apporter des recommandations pratiques et des exemples concrets directement applicables dans le travail de recherche

Modalités : cours, exercices, cas pratiques, indication des personnes ressources, utilisation des plans de gestion des données

Supports : disponibles [sur le site Science ouverte de l'UGA](#) (mai 2022)



1° journée :

- trouver des données : entrepôts et outils
 - licences juridiques liées aux données
 - principes FAIR
- organisation hiérarchique des fichiers, nommage, gestion de version, formats ouverts
- documenter des données et standards de métadonnées
- plan de gestion des données



2° journée :

- contexte national et européen de la science ouverte (enjeux, acteurs et outils)
- stocker et sécuriser des données
- aspects juridiques des clouds commerciaux
- trier et conserver des données (dont cadre légal)
- impact environnemental des données
- ressources et services mobilisables durant la thèse



1/2 journée sur les données à caractère personnel :

- cadre légal, dont RGPD
- principes légaux fondamentaux
- plan de gestion des données



1/2 journée sur la diffusion des données :

- entrepôts de données : dépôt et recherche
- datapapers
- ressources et services en appui pour la diffusion



**Ressources et accompagnement
pour les doctorants et les encadrants**



[Passeport pour la Science Ouverte](#)

Guide pratique à l'usage des doctorantes et des doctorants (2021)

[Portail OPIDor](#) et surtout [DMP OPIDor](#)

[DoraNum](#)

Nouvelles ressources à venir sur le site « [Ouvrir la science](#) »

Révision du passeport en cours (diffusion automne 2022)

Guide sur la gestion des données

Guide sur les codes et logiciels

Idées reçues sur la science ouverte

3 capsules vidéos

A voir aussi : Magali Civet, Delphine Quentin, & Claire Tignolet. (2021). [Bingo de la Science Ouverte](#)



- Quelle connaissance des outils et services locaux ou nationaux par les doctorants ?
 - Comment diffuser l'information sur ces services ?
- Quels éléments vous semblent nécessaires pour un plan de gestion des données adaptable au travail des doctorants ?



- Quels sont les besoins de formations :
 - autour de la science ouverte ?
 - autour de la gestion des données ?
 - autour des codes et logiciels ?
- Y a-t-il des besoins d'accompagnement spécifique ? Si oui :
 - sur quelles thématiques ?
 - selon quelles modalités ?
- *Exemple : si un travail de recherche se déroule sur plusieurs doctorats successifs, comment se passe la transmission des informations et des réalisations entre doctorants ?*



Y a-t-il besoin d'un accompagnement sur les questions juridiques liées à :

- La propriété intellectuelle ?
- Le respect du RGPD ?
- Les données ?
- Les codes et logiciels ?
- Les publications ?



- A quelles autres questions êtes vous confrontés ?
 - Recherche de données
 - Stockage des données
 - Utilisation des outils collaboratifs
 - Description des données
 - Modalités de diffusion des données
 - Science ouverte
- Quels services vous manquent pour accompagner les doctorants ?



- **Séminaire régulier** (annuel ou bisannuel) pour les directeurs de thèse
- **Groupe de travail** avec dans chaque école doctorale :
 - 1 ou 2 référents « directeurs de thèse »
 - 1 ou 2 référents doctorants→ Objectif : faire remonter les besoins, les questions, adapter l'accompagnement aux différentes disciplines
- **Recensement des questions** qui se posent sur la gestion des données pour les doctorants et des réponses apportées
- **Forum d'échange**, par exemple sur le site science ouverte de l'UGA
- **Journée annuelle de « rentrée »** sur la gestion des données pour les doctorants et les directeurs de thèses
 - Echange d'information

Autres idées ?