

Module doctorant B 10 : Comment bien gérer ses données pour faciliter son travail de doctorat ?
Enjeux et bonnes pratiques, BAPSO, CDGA-UGA, 2026

Lien du cours en ligne :

Fiche synthèse du chapitre 2 : Organiser ses données

1 Faire un Plan de Gestion de Données pour sa thèse

Le Plan de Gestion des Données (PGD) est un document qui permet de **regrouper tous les éléments concernant la gestion des données de recherche** d'un projet ou d'une thèse. Il est obligatoire dans le cadre d'un projet financé.

Il permet de réfléchir et de formaliser la gestion de vos données **à toutes les étapes de leur cycle de vie**. Vous serez donc amené **à le faire évoluer** au fur et à mesure du déroulement de votre thèse. C'est **une base de discussion avec vos encadrants**, votre équipe.

Les différentes questions à aborder dans un PGD :

- **Les informations générales** de la thèse (titre, résumé)
- **La description des données** : nature, type de données, volumétrie, production, collecte, réutilisation...

- **La documentation et la qualité des données :**
organisation, description des données...
- **Les aspects légaux et éthiques :** À qui appartiennent les données que vous produisez ou collectez ? Quelles démarches entreprendre si vous traitez des données personnelles ou sensibles ?
- **Le traitement et l'analyse des données :** nettoyage, croisement, traitement par lot...
- **Les modalités de stockage et de sauvegarde pendant la thèse :** stockage sur le serveur du laboratoire ? Quelles sauvegardes automatiques et à quelle fréquence ? Qui a accès aux données avec quel niveau de sécurité ?
- **Les modalités de diffusion et d'archivage des données à l'issue de la thèse :** choix de partage envisagés (entrepôts de données, publications)

2 Organiser ses données

Prenez le temps au départ de réfléchir à votre organisation ; restez **pragmatique et réaliste** (ce qui est utile pour vous) et **prenez en compte votre contexte de travail** (les pratiques de votre communauté).

2.1 Description

Décrire **vos jeux de données dès l'étape de collecte et de production** vous permettra de comprendre à quoi ils correspondent même 2 ans après leur production. Une description complète permettra un gain de temps pour la rédaction et pour la diffusion des jeux de données en fin de thèse.

Un **fichier Read-me** décrivant de manière complète la **nature** de vos données, leur **contexte de production** et la **méthode**

d'acquisition utilisée, leur **organisation** (arborescence et description des fichiers), les **caractéristiques techniques** (format et volume des fichiers, logiciel à utiliser pour les consulter), leur **potentiel de réutilisation** (par vous ou par d'autres).

2.2 Organisation en dossiers

Mettre en place une **arborescence de dossiers** selon une logique pragmatique (et pratique pour vous) vous permettra de **retrouver facilement et rapidement** n'importe quel contenu.

- **Construisez votre arborescence hiérarchisée**, à base de répertoires et sous-répertoires de façon logique, en suivant une organisation "en entonnoir", c'est à dire du plus général au plus particulier.
- **Décrivez** votre arborescence dans un fichier Read-Me.

2.3 Nommage des fichiers

Votre façon de nommer vos fichiers est complémentaire à l'organisation de vos dossiers.

Il est important d'écrire une **convention de nommage** de vos fichiers pour **documenter** la construction des **noms de vos répertoires et de vos fichiers** et pour en garder la trace.

Quelques recommandations :

- un nom court
- la même langue
- des majuscules ou des "tiret du bas" pour séparer les informations
- la date au format AAAA-MM-JJ

- Ne pas utiliser d'espace, ni de caractère spécial (accents, cédille, #, ?, etc.)

2.4 Gestion des versions

Pour suivre les changements de version, il existe **deux solutions**.

Utiliser la **convention de nommage** de vos fichiers pour préciser la **version et l'auteur** de la modification dans le nom du fichier.

Autre solution, utiliser un **gestionnaire de version**, quand plusieurs personnes sont susceptibles de modifier les fichiers. Si vous manipulez des documents en **texte brut**, il est conseillé d'utiliser le logiciel Git.

3 Stockage

Le **stockage** est l'enregistrement d'une information sur un support physique. Il existe différentes solutions présentant des avantages et des inconvénients (disque dur, serveur du laboratoire, cloud institutionnel).

Voir annexe 2 : [tableau comparatif des différents supports de stockage](#)

Privilégier les solutions académiques ou mises à disposition de votre laboratoire.

Il est nécessaire d'ajuster le lieu du stockage de la donnée, en fonction des besoins (par exemple, si on doit y accéder de manière fréquente, c'est à dire pour des données chaudes). Pour optimiser le temps de transfert, il est ainsi préférable de rapprocher les données chaudes de leur lieu de traitement. Les données froides peuvent par contre être éloignées sans difficulté.

3.1 Format

Le **Format** c'est la manière dont les données sont structurées, organisées et encodées sur le support physique.

La recommandation est de privilégier les formats ouverts, non propriétaires, ou dont l'usage est répandu dans sa communauté de recherche :

Voir la [liste indicative de formats ouverts et fermés](#) sur DoraNum, en annexe 3.

3.2 Volumétrie

Dès que le volume de données à traiter est limité par la mémoire vive de son propre ordinateur, il devient important. Dans tous les cas, à partir de quelques dizaines de TeraOctets (To), il s'agit de grosses volumétries. Dans ce cas, la circulation des données complexifie et ralentit les transferts des données par le réseau.

Quelques échelles :

- 1 Go = 1 000 Mo, 1 To = 1 000 Go, 1 Po = 1 000 To

Ces éléments de volume sont à prendre en compte pour le stockage des données.

4 Sauvegarde

La sauvegarde consiste à dupliquer des données pour les mettre en sécurité sur des supports de stockage différents (par exemple : faire régulièrement des copies des fichiers de son ordinateur portable sur un disque dur externe, sur le serveur du labo) :

L'objectif de la sauvegarde est de pouvoir facilement récupérer des données en cas de perte ou de mauvaise manipulation, c'est l'opération de restauration.

4.1 Sécurisation

Tout au long de votre thèse, il est fortement recommandé

- **d'appliquer la règle de la sauvegarde 3-2-1** (3 copies sur 2 supports différents, avec au moins 1 une copie à distance)

et de faire une **sauvegarde quotidienne** de son travail (automatique dans le cas de solutions académiques)

- De **sécuriser son poste de travail** (d'autant plus si vous traitez des données à caractère personnels ou si vous êtes en Zone à Régime Restrictif)
 - Via le chiffrement (obligatoire à l'UGA) et le verrouillage de votre session dès que vous vous éloignez de votre ordinateur
 - Via un mot de passe solide.

4.2 Usage de solutions académiques

Il est recommandé d'éviter les clouds commerciaux et d'utiliser les solutions académiques car elles offrent des garanties de sécurité, de fiabilité et elles sont de plus conformes aux exigences de respect du RGPD et des ZRR.

- A l'UGA : le CloudUGA, Nova, SUMMER ;
- Au niveau national : des services comme Sharedocs Humanum, Resana.

5 Décrire ses données pour les diffuser

Le fait d'avoir décrit vos données au fur et à mesure de leur collecte ou production vous permet de disposer d'informations fiables pour votre travail et de préparer la diffusion de vos données selon les principes FAIR.

Pour achever de mettre en œuvre ces principes, il est recommandé d'ajouter si vous ne l'avez pas fait au fur et à mesure :

- Une licence d'usage
- Des identifiants pérennes, qui permettront un accès durable et non ambigu à vos données

- DOI pour jeux de données
- ORCID pour les auteurs.

6 Conclusion

L'organisation de vos données au cours de votre thèse relève du bon sens. Elle est sous-tendue par l'application des principes FAIR pour vous permettre de vous y retrouver ; ces conseils et recommandations sont également pertinents pour la rédaction de votre thèse et au-delà dans votre travail de recherche.

La Cellule Codes Données Grenoble Alpes vous accompagne tout au long de votre thèse. Elle vous propose de vous accompagner pour la rédaction d'un plan de gestion des données. Vous pouvez également bénéficier de rendez-vous personnalisés à tout moment.

Vous pouvez nous envoyer le PGD simplifié pour relecture et accompagnement.

Adresse de contact : sos-data@univ-grenoble-alpes.fr

Vous pouvez également contacter le référent données de votre laboratoire.

Si vous êtes concernés par l'usage de données personnelles ou sensibles, il est indispensable de prendre rendez-vous avec le délégué à la protection des données le plus tôt possible. Son rôle est de vous accompagner dans le respect de la protection des données tout au long de votre recherche et de vous aider à mettre en place des solutions appropriées. Contact de la DPO de l'UGA : dpo@univ-grenoble-alpes.fr

7 Annexes

7.1 Annexe 1 : Ressources

Le [site web science ouverte UGA](#) propose de nombreuses ressources sur la gestion des données mais aussi sur les publications, les codes et logiciels de recherche, ainsi que sur la démarche science ouverte.

Le [Passeport science ouverte Données de la recherche](#) offre une excellente synthèse complémentaire de ce cours.

Le [site d'autoformation DoraNum](#) est également un réservoir précieux de ressources et de fiches pratiques.

Le [catalogue CatOpidor](#), qui cartographie les services français dédiés aux données scientifiques pour chaque étape du cycle de vie, recense plus de 400 services pour l'analyse des données (plateformes de calcul, outil de gestion des données, services d'accompagnement, entrepôts) et indique les disciplines concernées.









7.2 Annexe 2 : Tableau comparatifs des lieux de stockage

Support de stockage	Sécurité	Accès	Coût	Remarque d'utilisation
 Ordinateur professionnel	★★☆☆ Sujet au piratage informatique, aux détériorations et pannes	★★☆☆ Pas adapté au partage, nécessite l'utilisation d'un support externe ou d'Internet (mail, cloud...)	★★☆☆ Pas de coût supplémentaire ou coût peu important	- Pour un stockage temporaire - Nécessité de crypter les données confidentielles et sensibles
 Support externe	★☆☆☆ - Sujet au vol, à la perte du support - Durée de vie limitée (dégradation du matériel)	★★☆☆ Facilement transportable, il permet de transférer les données vers un autre ordinateur	★★☆☆ Pas de coût supplémentaire ou coût peu important	- Pour un stockage temporaire - Nécessité de crypter ou de sécuriser physiquement les données confidentielles et sensibles
 Serveur institutionnel	★★☆☆ Stockage fiable, durable et sécurisé (contre le vol, le piratage, les incendies...)	★★☆☆ La connexion au serveur institutionnel ne facilite pas le travail avec des personnes extérieures	★★☆☆ Coût assez important mais pas forcément répercuté sur l'utilisateur	- Pour un stockage plus pérenne - Adapté pour le stockage de données sensibles et des versions « stables » de vos données - Toutes les institutions ne proposent pas ce service
 Serveur Cloud	★★☆☆ On ne sait pas vraiment où sont stockées les données, ni ce qu'elles deviennent	★★☆☆ Permet un travail synchronisé avec toutes les personnes ayant été autorisées au partage	★★☆☆ Payant à partir d'une certaine limite de stockage	- Pour un partage avec des personnes externes à l'institution - Ne pas y mettre de données sensibles ou confidentielles - Pas de contrôle sur la procédure de sauvegarde des données

[Retour au corps du document.](#)

7.3 Annexe 3 : Liste indicative des formats ouverts et fermés [\(retour vers le corps du document\)](#)

Les FORMATS OUVERTS vs FERMÉS (propriétaires)

<ul style="list-style-type: none"> • ODT (OpenDocument Text) • RTF (Rich Text Format) • TXT (Text) 	TEXTE 	<ul style="list-style-type: none"> • DOC (Microsoft Word) • PAGES (Apple) • WP (WordPerfect)
<ul style="list-style-type: none"> • BMP (Windows BitMaP) • GIF (Graphics Interchange Format) • JPG (Joint Photographic Expert Group) • PNG (Portable Network Graphics) 	IMAGE 	<ul style="list-style-type: none"> • JP2 (Joint Photographic Expert Group 2000) • TIF (Tagged Image File Format)
<ul style="list-style-type: none"> • CSV (Comma-separated values) • ODS (OpenDocument SpreadSheet) 	TABLEUR 	<ul style="list-style-type: none"> • XLS (Microsoft Excel) • NUMBERS (Apple)
<ul style="list-style-type: none"> • MKV (Matroska) • NUT • OGM (OGG Media) 	VIDEO 	<ul style="list-style-type: none"> • AVI (Audio Video Interleave) • FLV (Flash Video) • MOV (QuickTime Movie) • MP4 (MPEG-4 Part 14) • QT (QuickTime Movie) • WMV (Windows Media Video)
<ul style="list-style-type: none"> • FLAC (Free Lossless Audio Codec) • MP3 (MPEG-1/2_Audio_Layer_III) • OGG • WAV (Waveform Audio File Format) 	AUDIO 	<ul style="list-style-type: none"> • AAC (Advanced Audio Coding) • RAM (Real Audio Metadata) • WMA (Windows Media Audio)
<ul style="list-style-type: none"> • ODP (OpenDocument Presentation) 	PRÉSENTATION 	<ul style="list-style-type: none"> • KEY (Keynote Mac) • PPT (Microsoft PowerPoint)
<ul style="list-style-type: none"> • 7z (Seven Zip) • TAR (Tape ARchiver) • ZIP 	ARCHIVAGE 	<ul style="list-style-type: none"> • ACE • RAR (Roshal ARchive)
<ul style="list-style-type: none"> • EPS (Encapsulated PostScript) • ODG (OpenDocument Graphics) • SVG (Scalable Vector Graphics) 	DESSIN 	<ul style="list-style-type: none"> • AI (Adobe Illustrator Artwork) • CDR (CorelDRAW) • FH (Adobe FreeHand)

Publié le 03/08/2023
Mis à jour le 11/12/2023

