

Plan de Gestion de Données

Recommandations à l'ANR

Jun 2019

Collège « Données »
Groupe de travail « PGD »

Plan de gestion de données - Recommandations à l'ANR

Comité pour la Science ouverte (CoSO) / collège "données" / groupe de travail PGD

L'ANR a sollicité le Comité pour la science ouverte (CoSO) pour bénéficier de recommandations dans la mise en place d'un plan de gestion des données pour les projets financés à partir de 2019, notamment sur le volet "évaluation". Pour répondre à cette demande le groupe de travail¹ « PGD » du collège « données » formule des propositions dans ce document.

Une attention particulière a été portée à la prise en compte des différents niveaux de maturité des communautés scientifiques dans la gestion des données et leurs éventuelles craintes face à la science ouverte.

1. La gestion des données de la recherche – Éléments généraux

La politique science ouverte de l'ANR fait l'objet d'une page web sur son site. Pour une plus grande lisibilité, un document dédié pourrait reprendre les principes clés de la politique de science ouverte de l'ANR concernant les données de la recherche, avec par exemple :

- Un encouragement aux bonnes pratiques de recherche, notamment :
 - la rédaction d'un plan de gestion des données ;
 - le respect progressif des principes FAIR² ;
 - le respect des principes d'intégrité scientifique (citation des données ré-utilisées etc.) ;
 - le dépôt des données dans des entrepôts « dignes de confiance » (disciplinaires, institutionnels, ou à défaut zenodo.org) ;
 - la possibilité d'un embargo dans l'exploitation des données à l'issue du projet ;
 - la publication de *data papers* ;
 - la réutilisation de données pré-existantes.
- Un rappel du cadre juridique : "Aussi ouvert que possible, aussi fermé que nécessaire" (données sensibles, données personnelles, protection des possibilités de valorisation économique...)
- Le rôle de l'accord de consortium pour préciser les modalités de valorisation des données produites au cours du projet, surtout en cas de partenariats avec des entreprises.

¹ Zoé, Ancion (ANR), Francis André (CNRS), Sarah Cadorel (Sciences Po), Romain Féret (Université de Lille), Odile Hologne (INRA), Kenneth Maussang (Université de Montpellier), Marine Moguen-Toursel (ENSEA), Véronique Stoll (Observatoire de Paris).

² Acronyme pour *Findable, Accessible, Interoperable, Reusable*, soit *Facile à trouver, Accessible, Interopérable, Réutilisable*. Le respect des principes FAIR n'entraîne pas une ouverture automatique des données, mais porte sur leur bonne gestion.

Recommandation n°1 :

Préciser dans un document dédié les éléments généraux soutenus par l'ANR concernant les données de la recherche.

Les PGD (Plan de gestion de données) ou DMP (*data management plan*) constituent des outils indispensables de bonne gestion des données. Évolutifs au cours des projets de recherche, ils permettent d'explicitier le cadre de la collecte et de l'organisation des données, mais aussi de leur accès, leur partage et leur réutilisation future, et ainsi d'anticiper les problèmes techniques ou juridiques. Ils constituent des outils d'accompagnement et d'incitation de bonnes pratiques de gestion, pouvant déboucher, si les conditions sont réunies, sur une ouverture des données.

Recommandation n°2 :

Privilégier un modèle de plan de gestion de données tel que celui proposé par Science Europe car il insiste sur les bonnes pratiques de gestion et partage.

2. Phasage et étapes de mise en œuvre

Une démarche progressive favoriserait une adhésion de la part des communautés et une meilleure adaptation à l'évolution des pratiques, tout en permettant le développement des services pour le stockage et la « FAIRisation » des données.

Recommandation n°3

Mettre en place une démarche pour l'adoption des plans de gestion pour aller progressivement vers l'ouverture :

- Étape 1 : Stimuler les bonnes pratiques de gestion de données
- Étape 2 : Prendre en compte la gestion des données dans l'évaluation des projets
- Étape 3 : Encourager et suivre l'ouverture

Étape 1 : Stimuler les bonnes pratiques de gestion des données**Recommandation n°4**

Faire une évaluation légère des PGD remis à 6 mois dans une optique de conseil aux porteurs de projets, et encourager la mise à jour ultérieure du PGD (rapport intermédiaire, fin de projet).

Étape 2 : Intégrer la science ouverte et la gestion des données dans la sélection des projets de recherche

Proposer la prise en compte des principes de gestion des données dès la phase de sélection des projets. L'évaluation de cette partie dédiée à la gestion des données pourrait porter par exemple sur la prise en compte de la gestion des données dans le programme de travail, la gestion des risques, la cohérence entre les objectifs et les ressources, la stratégie de dissémination (lien entre publications et données, la stratégie d'archivage et de diffusion des données)...

Recommandation n°5 :

Mentionner la gestion des données dans la trame de la proposition détaillée des appels à projets et ajouter un sous-critère d'évaluation *ad hoc* dans la grille à destination des évaluateurs.

Étape 3 : Encourager et suivre l'ouverture des données

Favoriser l'ouverture et la mise à disposition des données dans le respect du principe « aussi ouvert que possible aussi fermé que nécessaire ». Le plan de gestion des données devrait être rédigé en s'appuyant sur les principes FAIR. Par ailleurs, dans un souci d'accroissement de la visibilité et de la mise à disposition des jeux de données issus des projets financés par l'ANR, ceux-ci devraient être référencés (utilisation d'identifiants pérennes...). Ainsi, lors du dépôt d'un jeu de données dans un entrepôt, la mention d'un financement de l'ANR devrait être systématique, de même que l'identifiant du projet, à l'image des pratiques en matière de publication. L'impact scientifique du projet pourrait ainsi se mesurer aussi à partir d'indicateurs associés aux jeux de données, à l'instar des indicateurs bibliométriques.

Recommandation n°6 :

Encourager la publication de data papers (articles de données) qui permettent de décrire les jeux de données publiés, de favoriser leur réutilisation et d'être cités dans des articles scientifiques classiques.

Recommandation n°7 :

La version finale du plan de gestion des données devrait mentionner les actions concrètes prévues par les porteurs de projet pour ouvrir leurs données et faciliter leur accès, leur valorisation et leur réutilisation après la fin du projet.

Recommandation n°8 :

Dans les métadonnées décrivant les données, demander l'indication de l'identifiant du projet et du financement, de manière analogue aux publications (section '*funding*' d'un article).

3. Dispositifs d'accompagnement

La maturité des communautés scientifiques concernant la gestion et l'ouverture des données étant différentes, nous préconisons de mettre en place un accompagnement qui s'appuie sur des dispositifs existants au niveau national ou à celui des établissements.

Recommandation n°9 :

Créer une page web didactique qui renvoie vers les principales ressources (guides, outils pour élaborer un PGD, exemples de PGD déjà rédigés, tutoriels...), renvoyer vers des "*success stories*" (réutilisation) ou des "*sad stories*" (pertes de données).

4. Suivi

1 : mettre en place un dispositif de suivi

Mettre en place un dispositif de suivi, en prenant soin de distinguer les questions liées à la gestion des données de la recherche de celles liées à leur ouverture. Ce dispositif de suivi pourrait démarrer avec la mise en place d'une enquête « observatoire » auprès des porteurs de projets après la réception des PGD avec pour objectif :

- D'identifier les communautés scientifiques nécessitant un accompagnement plus soutenu en matière de gestion et d'ouverture des données de la recherche ;
- D'identifier les difficultés des porteurs de projets, notamment l'impact budgétaire associé et les moyens RH nécessaires ;
- D'évaluer l'évolution des pratiques dans les différentes communautés scientifiques ;
- De permettre à terme la construction d'indicateurs associés aux jeux de données financés (utilisation des données hors projet, citations etc.) et la construction d'instruments de mesure d'impact des projets.

Recommandation n°10

Mettre en place une enquête « observatoire » auprès des porteurs de projets financés pour appréhender les pratiques des communautés et leurs difficultés. Cette enquête pourrait être réalisée par un tiers

Recommandation n°11

Former des personnes à l'évaluation des plans de gestion de données.

2 : identifier les coûts associés à une bonne gestion des données de la recherche

Une bonne gestion des données de la recherche nécessite un investissement financier (infrastructures) et humain (temps de travail). La viabilité des projets retenus par l'ANR impliquera donc une bonne estimation de ces dépenses par le porteur.

Recommandation n°12 :

Identifier et lister les types de dépenses associées à la gestion des données qui sont éligibles au financement par l'ANR.

Recommandation n°13 :

Demander au porteur de préciser les coûts associés à la gestion des données du projet (incluant le stockage et la mise à disposition).

Le dispositif de suivi devrait permettre de construire des indicateurs quantitatifs à usage interne de l'ANR pour la construction d'instruments de mesure d'impact des projets financés. L'analyse des rapports financiers finaux pourrait permettre d'obtenir une bonne visibilité des budgets requis pour la production et la gestion des jeux de données pour chaque domaine disciplinaire.

Recommandation n°14 :

Définir des indicateurs quantitatifs et des instruments de mesure d'impact associés aux données ouvertes des projets financés.

5. Évaluation des plans de gestion de données

Dans la phase d'expérimentation, la commission européenne a évalué les plans de gestion de données de projets en faisant appel à des experts et en organisant une formation des « *project officers* ». L'évaluation est basée sur le modèle du plan de gestion de la commission. Elle consiste à indiquer si l'information est correctement, partiellement ou mal décrite et à formuler des recommandations pour chaque critère. Avec une telle approche, le travail d'analyse peut prendre jusqu'à 4 heures par PGD.

Recommandation n°15 :

Nous suggérons :

- La mise en place d'un « pilote » ;
- Une évaluation-conseil basée sur la trame du plan de gestion de données de Science Europe ;
- D'avoir une attente différente quant au niveau de description en fonction de la version du PGD (un PGD à 6 mois peut être imprécis sur les modalités de diffusion en fin de projet) ;
- De formuler des recommandations globales ou des alertes pour les responsables de projets (par exemple dans le cas où des projets gèrent des données personnelles ou des données sensibles sans mettre en œuvre les bonnes pratiques de gestion ou d'ouverture de ces données).

Critère note	0	1	2
Description des données et collecte ou réutilisation des données existantes			
Documentation et qualité des données			
Stockage et sauvegarde pendant le processus de recherche			
Exigences légales et éthiques, codes de conduite			
Partage des données et conservation à long terme			

Signification des scores :

0 : rubrique peu ou mal renseignée

1 : rubrique partiellement renseignée

2 : rubrique bien décrite

Bibliographie PGD³

Modèles

Références

1. Doorn, P., Science Europe. (2018). Science Europe Guidance. Presenting a Framework for Discipline-specific Research Data Management. http://www.scienceeurope.org/wp-content/uploads/2018/01/SE_Guidance_Document_RDMPs.pdf
2. Doorn, P., & Timmermann, M. (2018). Towards Domain Protocols for Research Data Management (IG Domain Repositories RDA 9th Plenary meeting Community-driven Research Data Management). Paper presented at the 9. Plenary meeting Community-driven Research Data Management, Barcelona. <https://www.rd-alliance.org/sites/default/files/attachment/RDA%20DRIG%20Domain%20Protocols%20V3%20Barcelona%20April%202017%20-%20DoornAerts.pptx>
3. European Research Council. ERC Data Management Plan template. (12/04/2017). Consulté à l'adresse <https://erc.europa.eu/content/erc-data-management-plan-template>
4. Reymonet, N., Moysan, M., Cartier, A., & Délémontez, R. (2018). Réaliser un plan de gestion de données « FAIR » : modèle. Consulté à l'adresse https://archivesic.ccsd.cnrs.fr/sic_01690547/document

A propos des futurs Machine-actionable PGD :

5. Miksa, T., Simms, S., Mietchen, D., & Jones, S. (2018). Ten simple rules for machine-actionable data management (preprint). *PLoS ONE*. <http://doi.org/10.5281/zenodo.1172673> et <https://zenodo.org/record/1172673#.Wt4HIZcuBaQ>
6. Miksa, T., Neish, P., Walk, P., & Rauber, A. (2018). Defining requirements for machine-actionable Data Management Plans [Preprint]. <https://zenodo.org/record/1266211>
7. Simms S, Jones S, Mietchen D, Miksa T (2017) Machine-actionable data management plans (maDMPs). *Research Ideas and Outcomes* 3: e13086. <https://doi.org/10.3897/rio.3.e13086>

Sites

5. Modèles existants (listes) – Financeurs
site DMPTool : https://dmptool.org/public_templates
site DMPonline : https://dmponline.dcc.ac.uk/public_templates

³ Bibliographie initiée par le pôle “Données de la recherche” de l’Inra <https://www6.inra.fr/datapartage/Gerer/Rediger-un-plan-de-gestion>

Recommandations

Références

1. Science Europe. (Décembre 2018). PRACTICAL GUIDE TO THE INTERNATIONAL ALIGNMENT OF RESEARCH DATA MANAGEMENT https://www.scienceeurope.org/wp-content/uploads/2018/12/SE_RDM_Practical_Guide_Final.pdf
 2. ERC. Open Research Data and Data Management Plans Information for ERC grantees. (2018, février 23). 24/04/2018 <https://erc.europa.eu/content/open-research-data-and-data-management-plans-information-erc-grantees>
 3. Hooft, R. Elsevier Publishing Campus. (s. d.). How to create a good data management plan. Consulté à l'adresse <https://www.elsevier.com/authors-update/story/publishing-trends/how-to-create-a-good-data-management-plan>
 4. Landelijk Coördinatiepunt Research Data Management (LCRDM). (s. d.). 10 tips for writing a Data Management Plan. Consulté à l'adresse https://www.edugroepen.nl/sites/RDM_platform/Shared%20Documents/Bij%20de%20WG%20Onderzoeksondersteuning%20en%20advies/LCRDM%2010%20tips%20for%20writing%20a%20DMP%20no%20branding.pdf
- Research Data Alliance. (13:04:53 UTC). *OpenAIRE and Eudat services and tools to support FAIR DMP implementation...* Données & analyses. Consulté à l'adresse <https://fr.slideshare.net/ResearchDataAlliance/openaire-and-eudat-services-and-tools-to-support-fair-dmp-implementation-68900290>
6. Reymonet, N., Moysan, M., Cartier, A., & Délémontez, R. (2018). Réaliser un plan de gestion de données « FAIR » : modèle. Consulté à l'adresse https://archivesic.ccsd.cnrs.fr/sic_01690547/document

Sites

1. **CIRAD** Se familiariser avec les plans de gestion de données de la recherche <https://coop-ist.cirad.fr/gestion-de-l-information/gerer-les-donnees-de-la-recherche/se-familiariser-avec-les-plans-de-gestion-de-donnees-de-la-recherche/3-exemple-de-trame-d-un-plan-de-gestion-de-donnees>
2. **DCC** Example DMPs and guidance <http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples>
3. **ESRC** - Research data Policy <https://esrc.ukri.org/funding/guidance-for-grant-holders/research-data-policy/>
4. **INRA** Plan de gestion des données <https://www6.inra.fr/datapartage/Gerer/Plan-de-gestion>
5. **IRSTEA** DMP Plan de gestion de données <https://donnees-recherche.irstea.fr/dmppgd-plan-de-gestion-de-donnees/>
6. **RIO** Section=[Data Management Plans] [https://riojournal.com/browse_journal_articles.php?form_name=filter_articles&sortby=0&journal_id=17&search_in =0§ion_type\[\]=231](https://riojournal.com/browse_journal_articles.php?form_name=filter_articles&sortby=0&journal_id=17&search_in =0§ion_type[]=231)

7. **Sciences Po** Libguides - Données de la recherche <https://sciencespo.libguides.com/donnees-de-la-recherche/dmp>

Evaluation

Références

1. Carlson, J., Wells Parham, S., Hswe, P., & Whitmire, A. (2016). Using data management plans to explore variability in research data management practices across domains. Paper presented at the IDCC 2016, Amsterdam (NLD). <http://slideplayer.com/slide/10989016/>
2. Grootveld, M., & Leenarts, E. (2018). Why is this a good DMP ? Paper presented at the TUD Seminar "Towards cultural change in data management - data stewardship in practice". https://pure.knaw.nl/portal/files/6616988/20180524_Why_is_this_a_good_DMP_public.pdf
3. Grootveld, M., & Van Selm, M. (2017). Report : survey of DMP reviewer experiences. Retrieved from https://www.edugroepen.nl/sites/RDM_platform/Shared%20Documents/Bij%20de%20WG%20Onderzoeksondersteuning%20en%20advies/Report%20about%20LCRDM%20DMP%20review%20survey%20-%2022%20June%202017.pdf
4. Grootveld, M., Leenarts, E., Jones, S., Hermans, E., & Fankhauser, E. (2018, janvier 9). OpenAIRE and FAIR Data Expert Group survey about Horizon 2020 template for Data Management Plans. Zenodo. Consulté à l'adresse <https://zenodo.org/record/1120245>
5. Jones, S. (2017). Developing and Reviewing Data Management Plans. Paper presented at the DCC - Supporting Open Research, Amsterdam (NLD). www.dcc.ac.uk/webfm_send/2384
6. Jones, S. (2018). What you need to know about DMPs. Paper presented at the FOSTER & OpenAIRE webinar (22/10/2018). <https://doi.org/10.5281/zenodo.1461601>
7. Jones, S., Leenarts, E., Grootveld, M., Fankhauser, E., & Hermans, E. (2018). OpenAire and the FAIR Data Expert Group. Report on the results of the survey about Horizon 2020 template for Data Management Plans. Consulté à l'adresse : <https://zenodo.org/record/1120245#.Wl2VAzciGM8>
8. Van Loon, J. E., Akers, K. G., Hudson, C., & Sarkozy, A. (2017). Quality evaluation of data management plans at a research university. IFLA Journal, 43(1), 98-104. doi : [10.1177/0340035216682041](https://doi.org/10.1177/0340035216682041)
9. Whitmire, A., Carlson, J., Hswe, P., Wells Parham, S., & Westra, B. (2016). Analysing DMPs to inform research data services. Lessons from the DART Project. Paper presented at the IDCC 2016, Amsterdam (NLD). <https://fr.slideshare.net/amandawhitmire/idcc-workshop-analysing-dmps-to-inform-and-empower-academic-librarians-in-providing-research-data-support-lessons-from-the-dart-project>
10. Whitmire, A., Westra, B., Hswe, P., Carlson, J., & Parham, S. W. (2016). Using Data Management Plans to Explore Variability in Research Data Management Practices Across Domains. International Journal of Digital Curation, 11(1), 53-67. doi: [10.2218/ijdc.v11i1.423](https://doi.org/10.2218/ijdc.v11i1.423)

Sites

1. **The DART Project**: using data management plans as a research tool. <https://osf.io/kh2y6/>
2. Site **DCC** – Exigences des financeurs : <http://www.dcc.ac.uk/resources/data-management-plans/funders-requirements>

3. **University of Sheffield** - Data Management Plan Compliance Rubrics (aide à l'auto-évaluation)
<http://www.dcc.ac.uk/resources/data-management-plans/funders-requirements>
4. Exemples de grilles d'évaluation : **Wellcome Trust**
https://zenodo.org/record/257650/files/Wellcome_Rubric_v2.1.pdf, **EPSRC**
<https://zenodo.org/record/247087#.WtTGRJcuBaQ>
5. **LIBER** - DMP Cat (projet de catalogue de DMP avec évaluation publiée - seuls 7 en ligne à ce jour)
<https://libereurope.eu/dmpcatalogue>