



HAL
open science

Les plans de gestion de données (PGD)

Sylvie Cocaud, Dominique L'Hostis, Pascal Aventurier

► **To cite this version:**

Sylvie Cocaud, Dominique L'Hostis, Pascal Aventurier. Les plans de gestion de données (PGD). Cahier des Techniques de l'INRA, 2018, N° Spécial: Données de la recherche, 9 p. hal-04693775

HAL Id: hal-04693775

<https://hal.inrae.fr/hal-04693775v1>

Submitted on 11 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Les plans de gestion de données (PGD)

Sylvie Cocaud¹, Dominique L'Hostis², Pascal Aventurier³

Résumé. Les acteurs de la recherche - organismes et universités, financeurs publics et privés, éditeurs scientifiques - renforcent actuellement leur politique en faveur de la science ouverte et de l'ouverture des données tout en promouvant le respect des principes FAIR pour les données qu'ils produisent ou financent. Le plan de gestion de données (PGD ou DMP : Data Management Plan) constitue un élément clé pour atteindre ces objectifs et sa rédaction, à partir des nombreux modèles désormais disponibles, devient un exercice recommandé, voire obligatoire pour les coordonnateurs de projets financés sur fonds publics. Des outils logiciels, intégrant des aides à la rédaction sont déployés pour faciliter leur création. Celle-ci devrait être facilitée à l'avenir, lorsque l'interopérabilité de ces outils avec d'autres systèmes informatiques autorisera des échanges de contenus. Des enquêtes montrent encore beaucoup de méconnaissance et de frilosité au sein de la communauté scientifique, mais un changement s'amorce, soutenu par les services d'appui à la recherche, parfois épaulés de scientifiques qui acceptent de jouer un rôle de médiation et de conseil favorisant l'adoption de bonnes pratiques au sein de leur propre communauté.

Mots clés : plan de gestion des données; plan de gestion de logiciel; plan de gestion de projet; plan de gestion de structure ; DMP-OPIDoR ; démarche FAIR ; PGD de projet; PGD de structure

Introduction

La politique d'ouverture des données, largement soutenue par les différents acteurs de la recherche (financeurs, éditeurs, acteurs institutionnels et politiques) s'accompagne d'une forte incitation à la mise en œuvre de bonnes pratiques, compatibles avec certains principes définis au plan européen, les FAIR Guiding Principles (<http://doi.org/10.1038/sdata.2016.18>) pour gérer et piloter les données aux différentes étapes de leur cycle de vie. Parfois perçue par les scientifiques comme une contrainte administrative, la rédaction des plans de gestion est pourtant le meilleur moyen de bien planifier cette gestion et un élément clé pour produire des données FAIR.

La première partie de cet article donne des éléments de contexte sur le développement des PGD et leur usage au plan national et international. La seconde présente les modèles créés par différents acteurs de la recherche, et spécifiquement les modèles développés à l'Inra. La dernière partie présente quelques informations issues de retours d'expérience de rédaction et d'évaluation des plans puis évoque les perspectives d'évolution de ces plans.

¹ INRA, Services déconcentrés d'appui à la recherche Grand Est-Nancy/Colmar, F-54280 Champenoux ; sylvie.cocaud@inra.fr

² INRA, Services déconcentrés d'appui à la recherche Pays de la Loire, F- 44316 Nantes Cedex 3 ; dominique.lhostis@inra.fr

³ INRA, Services déconcentrés d'appui à la recherche Provence-Alpes-Côte d'Azur, F- 84914 Avignon Cedex 9 (adresse actuelle : IRD, Centre de documentation, F-34394 Montpellier Cedex 5)

Le PGD : un document fédérateur

Rédiger un plan de gestion des données, une nouvelle contrainte pour le chercheur ?

Dès les années 1960 certains domaines, comme la recherche spatiale ou les sciences de l'environnement et de l'atmosphère, confrontés à des processus de gestion de données complexes et coûteux reposant souvent sur des collaborations internationales, ont utilisé les plans de gestion pour les besoins propres des projets, en particulier pour contrôler les étapes de collecte, de traitement et d'analyse des données.

A partir des années 2000, l'explosion du volume de données produites et les bénéfices potentiels de l'open data incitent les financeurs à recommander aux porteurs de projets d'accompagner leur demande de financement d'un plan de gestion des données. La National Science Foundation (US NSF) est ainsi le premier financeur à demander à ce qu'un plan de gestion des données qui seront préservées sur le long terme soit fourni avec la demande de financement d'un projet : *"The NSF should require that research proposals for activities that will generate digital data, especially long-lived data, should state such intentions in the proposal so that peer reviewers can evaluate a proposed data management plan"* (National Science Foundation, 2005). Cette recommandation deviendra une exigence pour tous les projets financés par la NSF à partir de 2011. La NSF a été rapidement suivie dans cette démarche par les organismes de financement de la recherche au Royaume-Uni, pour s'étendre à des pays de plus en plus nombreux (figure 1).



Figure 1. Financeurs de la recherche exigeant un plan de gestion des données (S. Cocaud, 2019)

Les organismes de recherche, les institutions, les universités s'engagent progressivement dans la voie de l'Open Science. La rédaction des plans de gestion apparaît comme un des éléments importants de la mise en œuvre de leur « Data Policy » (voir au Royaume-Uni : UK Institutional data policies⁴ et en France : Cartier A., Moysan M. et Reymonet N. (2015)).

⁴ UK Institutional data policies <http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies>

Premier jalon dans une démarche FAIR

Parfois ressenti par les chercheurs comme une contrainte externe, imposée par le financeur, le PGD est pourtant un élément clé pour mettre en place, ou conforter, les bonnes pratiques de gestion des données qui garantiront leur possible réutilisation dans un futur proche ou lointain. C'est pourquoi il est naturellement associé à la mise en place de la démarche FAIR⁵ et utilisé de plus en plus dans d'autres contextes que celui des projets financés.

De nombreuses universités au Royaume-Uni, aux Etats-Unis, aux Pays-Bas et en Allemagne proposent des formations à la rédaction de PGD aux doctorants afin qu'ils apprennent à bien gérer et valoriser les données produites au cours de leur thèse. En France, l'université de Lille 3 développe aussi cette approche (Chaudiron et al., 2015).

Les PGD sont également utilisés par les unités et les infrastructures de recherche (plateformes d'analyse, observatoires...) qui peuvent s'appuyer sur ce document pour planifier la gestion et la valorisation des données qu'ils produisent, décrire leurs prestations ou accompagner les utilisateurs dans la rédaction d'un PGD de projet. A l'Inra, à la demande de scientifiques, un groupe de travail a ainsi été créé pour mettre en place un modèle de PGD adapté à des structures⁶.

Par ailleurs, un plan de gestion de logiciel (PGL) ou software management plan centré sur le code ou le logiciel peut être complémentaire d'un plan de gestion de données de projet, les deux plans se référant alors mutuellement. Il existe des modèles spécifiques pour les PGL (The Software Sustainability Institute, 2018) et le projet PRESOFT⁷ de l'IN2P3 (CNRS) s'inscrit dans cette démarche.

Le cycle de la recherche et le processus de publication des données apparaissent de plus en plus intégrés. Au sein du projet de recherche, le plan de gestion oblige à anticiper toutes les étapes conduisant de la collecte à la gestion, puis au traitement et à la conservation des données. Il permet ainsi d'anticiper le devenir des différents jeux de données, de leur exploitation et leur valorisation, par exemple sous forme de data articles ou data papers⁸.

Les différents modèles de PGD

Le plan de gestion de données (PGD) associé à un projet rassemble dans un document unique les informations concernant les données et leur cycle de vie. Il décrit comment ces données seront obtenues, organisées, traitées, stockées, sécurisées, préservées et partagées au cours et à l'issue du projet. Il aide ainsi les partenaires à s'interroger, dès le montage du projet, sur les risques liés à la gestion des données, à identifier les responsabilités et rôles de chacun dans cette gestion, et à planifier les ressources nécessaires pour produire des données de qualité, fiables, qui seront disponibles sur le long terme et compréhensibles, facilitant ainsi une future réutilisation.

Mis à jour autant que nécessaire au cours du projet, le PGD en constitue un des livrables et sa version finale peut être publiée ou déposée dans un entrepôt avec les données qu'il décrit : *"The content in DMPs must be put to good use so they become a central hub of information on FAIR Digital Objects, interlinking ecosystem components"* (European Commission, 2018).

Financeurs et organismes de recherche proposent de plus en plus leurs propres modèles et guides pour faciliter la rédaction des PGD⁹.

⁵ F1000. (s. d.). Your go-to guide to making your data Findable, Accessible, Interoperable, and Reusable (FAIR). https://f1000.com/resources/FAIR_Open_GettingStarted.pdf

⁶ INRA - Trame Structure sur DMP OPIDoR. https://dmp.opidor.fr/template_export/1528925823.pdf

⁷ <http://www.france-grilles.fr/category/presoft/>

⁸ Publier un data paper : <https://www6.inra.fr/datapartage/Partager-Publier/Publier-un-Data-Paper>

⁹ Site Digital Curation Centre : Funders' data plan requirements. <http://www.dcc.ac.uk/resources/data-management-plans/funders-requirements>.

En parallèle, des démarches au niveau international se mettent en place pour trouver un noyau commun aux modèles de PGD afin d'en simplifier la rédaction, en particulier dans les projets multi partenaires, et de mieux exploiter les PGD rédigés. Science Europe¹⁰ et le Netherland Organisation for Scientific Research (NWO)¹¹ ont ainsi identifié six sujets qui devraient être présents dans tout modèle de PGD (Figure 2).

- 1. Data description and collection or re-use of existing data**
 - a. How will new data be collected or produced and/or how will existing data be re-used?
 - b. What data (for example the kinds, formats, and volumes) will be collected or produced?
- 2. Documentation and data quality**
 - a. What metadata and documentation (for example the methodology of data collection and way of organizing data) will accompany data?
 - b. What data quality control measures will be used?
- 3. Storage and backup during the research process**
 - a. How will data and metadata be stored and backed up during the research process?
 - b. How will data security and protection of sensitive data be taken care of during the research?
- 4. Legal and ethical requirements, codes of conduct**
 - a. If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?
 - b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?
 - c. How will possible ethical issues be taken into account, and codes of conduct followed?
- 5. Data sharing and long-term preservation**
 - a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?
 - b. How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?
 - c. What methods or software tools will be needed to access and use the data?
 - d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?
- 6. Data management responsibilities and resources**
 - a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?
 - b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

Figure 2. Core requirements form DMPs (Science Europe, 2018)

Certaines informations requises par le PGD peuvent être reprises de documents existants : documents qualité, accord de consortium... La rédaction d'un PGD mobilise néanmoins des compétences variées et il peut être utile dès leur création de faire appel à des collaborateurs : scientifiques, informaticiens, ingénieurs partenariat-projet, documentalistes...

¹⁰ Science Europe est une association basée à Bruxelles regroupant 52 agences de financement et organismes de recherche issus de 27 pays européens, parmi lesquels l'Agence nationale de la recherche.

¹¹ NWO est une organisation néerlandaise financeur de la recherche scientifique aux Pays-Bas.

La Commission Européenne et les Plans de Gestion de Données

Depuis 2017, les projets financés par la Commission Européenne dans le cadre du programme Horizon 2020 participent par défaut à l'action pilote "libre accès aux données de la recherche" qui vise à rendre accessible au plus grand nombre les données de recherche générées par ces projets. Les responsables des projets doivent rédiger un PGD indiquant les données qui seront librement accessibles à l'issue du projet (à minima les données et métadonnées nécessaires à la validation des publications), toute exception à l'ouverture des données devant être justifiée. La phrase « *as open as possible, as closed as necessary* » résume clairement la politique européenne en matière d'open data.



Figure 3. Infographie résumant la politique "Open Data" de la commission européenne (extrait de : *Open research data in Horizon2020*¹²)

La Commission européenne fournit sans toutefois l'imposer un modèle de PGD pour les projets organisé autour des quatre principes FAIR ([Guidelines on FAIR Data Management in Horizon 2020](#) v.3, 26 July 2016).

L'ANR et les Plans de Gestion de Données

En France, l'ANR demande également depuis début 2019 la rédaction d'un plan de gestion pour les projets qu'elle finance ([Plan d'action ANR 2019](#)). Le renforcement de la politique de l'agence en matière de données de recherche suit la publication en juillet 2018 du [Plan National pour la Science Ouverte](#) dans lequel le ministère de la Recherche de l'Enseignement Supérieur et de l'Innovation présente ses lignes directrices pour « *Structurer et ouvrir les données de la recherche* » en France. Le respect des principes FAIR est recommandé, l'ouverture des données encouragée et la généralisation des plans de gestion de données annoncée. L'ANR proposera courant 2019 son propre modèle de plan de gestion.

Les PGD à l'Inra

L'Inra soutient l'ouverture des données de recherche et dans le cadre de sa politique Open Science, recommande dès 2013 la rédaction d'un PGD dans tout projet donnant lieu à une production de données et soumis à une demande de subvention¹³. Par ailleurs, la préservation et l'archivage des données hébergées dans les centres de

¹² https://ec.europa.eu/research/press/2016/pdf/opendata-infographic_072016.pdf

¹³ Note de cadrage du Comité Directeur des Systèmes d'Information : Principes en matière de partage des données de la recherche. 2013. <https://intranet6.inra.fr/systemes-information/content/download/4045/39055/file/CDSI-communication-partage-donneesvd.pdf>

données gérés par l'Inra ou ses prestataires doivent suivre les indications données dans les PGD qui accompagnent ces données¹⁴.

Un groupe de travail sur les plans de gestion, constitué dans le cadre du chantier "Data partage" de l'Inra, a élaboré un modèle générique adapté aux projets et intégrant les recommandations du programme H2020. Ce modèle comporte 53 questions réunies en 9 chapitres, accompagnées de recommandations :

1. Informations sur le PGD
2. Informations sur le projet
3. Présentation succincte des données
4. Droits de propriété intellectuelle
5. Confidentialité des données
6. Partage des données à l'issue du projet
7. Description et organisation des données
8. Stockage et sécurité des données au cours du projet
9. Archivage et conservation des données après la fin du projet

Cette trame est disponible sur le site "Gestion et Partage des données scientifiques"¹⁵ ainsi que sur DMP OPIDoR¹⁶, outil de saisie et d'aide à la rédaction de plans de gestion mis en place par l'Inist-CNRS pour l'Enseignement Supérieur et la Recherche (identification via son Ldap Inra après création d'un compte).

L'administration de ces différentes trames de plan de gestion de données pour les projets ou les structures est actuellement assurée par la DIST (pôle Digitalist) qui propose également un service d'accompagnement (aide à la rédaction et relecture de plans) ainsi qu'une offre de formation spécifique.

Evolution et perspectives

L'évaluation des plans de gestion

Les processus de relecture, comme les grilles de rédaction, varient selon les financeurs.

La Commission Européenne, par exemple, fournit des guides à l'attention des relecteurs des PGD des projets qu'elle finance¹⁷. La lecture du plan doit leur permettre de comprendre comment la question des données et de leur potentiel de réutilisation est prise en compte par les partenaires du projet, et d'identifier précisément la nature des données produites ou utilisées, celles qui seront ouvertes et comment, celles qui ne le seront pas et pourquoi. Les relecteurs doivent également pouvoir accéder à l'ensemble des documents cités dans le plan de gestion.

Des grilles génériques, comme celles mises au point dans le cadre du projet DART¹⁸ ou fournies par les financeurs¹⁹, permettent une évaluation de premier niveau des PGD rédigés. Diffusées sur Internet, elles peuvent aussi être utilisées par les rédacteurs pour vérifier la qualité de leur PGD et l'auto-évaluer au regard des attentes spécifiques de leur institution ou leur financeur.

¹⁴ Charte pour le libre accès aux publications et aux données. Paris, FRA : Editions INRA. DOI : 10.15454/1.485854076583696E12

¹⁵ Trame type d'un PGD Inra <https://www6.inra.fr/datapartage/Media/Fichier/PGD/PGD-trame-type-v1>

¹⁶ INRA - Trame générique projet https://dmp.opidor.fr/template_export/1988740616.docx

¹⁷ Open Research DMP Assessment GRID https://drive.google.com/file/d/10cgiPgbK24FFP9WC_4oMV1J-UFGDTFGZ/view

¹⁸ The DART Project: using data management plans as a research tool. <https://osf.io/kh2y6/>

¹⁹ Checklists to evaluate data management plans for compliance with the following funder's requirements : Site The University of Sheffield <https://www.sheffield.ac.uk/library/rdm/dmp>

Vers des PGD exploitables par les machines

Au sein de la [Research Data Alliance](#), plusieurs groupes de travail s'intéressent aux plans de gestion de données. Un des axes de réflexion concerne le développement de « Machine-Actionable DMPs », des plans de gestion exploitables par machines, interopérables avec les différents systèmes, outils ou services engagés dans un processus de recherche. La rédaction d'un PGD pourrait ainsi être semi-automatisée, facilitant les échanges entre les différentes parties prenantes d'un projet (chercheur, administrateur d'entrepôt de données, financeur ...)

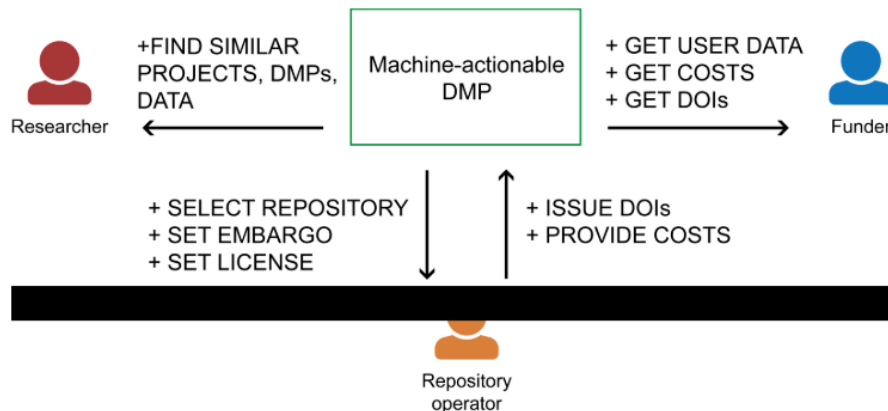


Figure 4. Echanges de contenus envisagés via les "Machine actionable DMP" (extrait de Miksa et al, 2019.)

Conclusion

Bien rédiger le PGD : un changement culturel plus que technologique...

Les enquêtes réalisées auprès de diverses communautés scientifiques montrent encore une assez forte méconnaissance des plans de gestion de données (O'Carroll et al., 2017), et plus largement de la démarche, des outils et des bonnes pratiques liés à l'Open Science. Trop fréquemment vécu comme une contrainte administrative, le PGD n'a souvent pas la qualité requise pour répondre à ses objectifs et aider réellement à la réutilisation des données (Smale et al., 2018).

Des services d'appui et de formation se mettent en place progressivement dans les institutions, souvent initiés et pilotés par les services en IST. L'implication de scientifiques sensibilisés et spécifiquement formés à cette problématique comme personnes ressources, relais et promoteurs des bonnes pratiques de gestion des données auprès des scientifiques, est indispensable pour aller plus loin. C'est le rôle que jouent, par exemple, les "Data Champions" à l'université de Cambridge²⁰ ou les "Data Stewards" à l'université de Delft²¹.

²⁰ Data champions <https://www.data.cam.ac.uk/intro-data-champions>

²¹ Data Stewardship <https://www.tudelft.nl/en/library/current-topics/research-data-management/research-data-management/data-stewardship/>

Pour se familiariser avec les PGD

- le site d'information Inra « Gestion et partage des données scientifiques » : <https://www.inra.fr/datapartage/>



- le site d'autoformation Doranum <http://doranum.fr/plan-de-gestion-de-donnees-dmp> propose un ensemble de ressources.
- Les formations en classe virtuelle pour découvrir les plans de gestion (Infodoc en ligne) : consulter le calendrier sur le site <http://ist.inra.fr/>

Références bibliographiques

Cartier A., Moysan M., Reymonet N. (2015). Construire des outils pour la gestion des données de la recherche dans une communauté d'universités. Journée sur les données de la recherche, Paris, France. [hal-01138663](https://hal.archives-ouvertes.fr/hal-01138663)

Chaudiron S., Maignant C., Schöpfel J., Westeel I. (2015). Les données de la recherche dans les thèses de doctorat - Livre blanc. Université de Lille 3. [hal-01192930](https://hal.archives-ouvertes.fr/hal-01192930)

European Commission. (2018). Turning FAIR into reality. Final Report and Action Plan from the European Commission Expert Group on FAIR Data. Consulté à l'adresse https://ec.europa.eu/info/publications/turning-fair-reality_en

Miksa, T., Simms, S., Mietchen, D., & Jones, S. (2019). Ten principles for machine-actionable data management plans. PLoS Comput Biol, 15(3), e1006750. doi:[10.1371/journal.pcbi.1006750](https://doi.org/10.1371/journal.pcbi.1006750)

National Science Foundation (2005). Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century (No. NSB-05-40) (p. 87 p.). Consulté à l'adresse <https://www.nsf.gov/pubs/2005/nsb0540/>

O'Carroll, C., Kamerlin, C. L., Brennan, N., Hyllsteh, B., Kohl, U., O'Neill, G., & Van den Berg, R. (2017). Providing researchers with the skills and competencies they need to practise Open Science. Report of the working Group of Education and Skills under Open Science: European Commission The Working Group on Education and Skills under Open Science. <http://doi.org/10.2777/121253>

Science Europe. (2018). Practical guide to the international alignment of research data management (p. 36 p.). Consulté à l'adresse

https://www.scienceeurope.org/wp-content/uploads/2018/12/SE_RDM_Practical_Guide_Final.pdf

Smale, N., Unsworth, K. J., Denyer, G., & Barr, D. P. (2018). The History, Advocacy and Efficacy of Data Management Plans. bioRxiv. <http://dx.doi.org/10.1101/443499>

The Software Sustainability Institute. (2018). Checklist for a Software Management Plan (Version 1.0). Zenodo. <http://doi.org/10.5281/zenodo.2159713>