



HAL
open science

Recommandations issues d'une analyse comparative des politiques de gouvernance des données (WP1) : Harmoniser les stratégies de politiques de gouvernance des données de la recherche (version 2)

S. Fortuno, Pascal Aventurier, D. Szabo, Luc Decker

► To cite this version:

S. Fortuno, Pascal Aventurier, D. Szabo, Luc Decker. Recommandations issues d'une analyse comparative des politiques de gouvernance des données (WP1) : Harmoniser les stratégies de politiques de gouvernance des données de la recherche (version 2). [Rapport de recherche] [s.n.]. 2023, 29 p. multigr. hal-04144546

HAL Id: hal-04144546

<https://hal.science/hal-04144546>

Submitted on 30 Jul 2023

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 4.0 International License



Recommandations issues d'une analyse comparative des politiques de gouvernance des données (WP1)

*Harmoniser les stratégies de politiques de
gouvernance des données de la recherche*

DOI : <https://doi.org/10.5281/zenodo.7507144>

Auteurs : Sophie Fortuno (Dgdrs - Mission Science ouverte et Données de la recherche, CIRAD), Pascal Aventurier (service Information Scientifique et Technique/ MCST, IRD), Dimitri Szabo (DipSO, INRAE), Luc Decker (service Information Scientifique et Technique/ MCST, IRD)



Sommaire

| | |
|---|----|
| I. Introduction | 3 |
| II. Périmètre du work package 1 du projet ANR-BRIDGE..... | 3 |
| 1. Définitions..... | 3 |
| 2. Politique de science ouverte, politique des données de la recherche..... | 4 |
| III. Démarche du work package 1..... | 5 |
| 1. Quel cadre commun adopter ?..... | 7 |
| 2. Méthode d'analyse du cadre de référence <i>RDA Data Policy</i> | 9 |
| IV. Cadre commun pour une politique des données de la recherche | 12 |
| 1. Définitions des types de données (portée de la politique)..... | 12 |
| 2. Les types de données à l'origine d'exceptions..... | 14 |
| 3. Données sous embargo..... | 15 |
| 4. Documents supplémentaires (<i>supplementary materials</i>)..... | 15 |
| 5. Entrepôts de données | 16 |
| 6. Citation des données | 18 |
| 7. Licences des données..... | 19 |
| 8. Soutien aux chercheurs auteurs | 20 |
| 9. Déclarations de disponibilité des données (<i>Data Availability Statement</i>) | 21 |
| 10. Formats et normes de (méta)données | 22 |
| 11. Partage obligatoire de types de données spécifiques..... | 23 |
| 12. Partage obligatoire de données | 23 |
| 13. Examen des données par les pairs..... | 24 |
| 14. Plan de Gestion de Données | 25 |
| V. Autres éléments pour appuyer une politique de données d'un établissement de recherche | 26 |
| 15. Légal et réglementaire | 26 |
| 16. Intégrité scientifique | 27 |
| VI. Conclusions et perspectives..... | 28 |

I. Introduction

Préambule

En France, la structuration de la recherche publique en Unités Mixtes de Recherche (UMR) nous amène tous les jours à produire des contenus numériques de manière conjointe. Pour cela, la coordination des politiques de partage et de libre accès entre instituts devient primordiale si l'on souhaite une mise en œuvre cohérente et efficace de la science ouverte à l'échelle nationale mais également internationale.

En tirant parti du contexte de partenariat favorable entre l'IRD, l'INRAE et le Cirad, le projet ANR-BRIDGE a pour ambition de démontrer comment, en activant trois leviers (politique, communautaire et technique), nous pouvons promouvoir la mise en œuvre des principes FAIR. Dans cet objectif, le projet ANR-BRIDGE a réuni des personnes ayant des expertises en matière de gouvernance des données, technique et scientifique dans différents champs disciplinaires.

Pour atteindre ce but, le projet ANR-BRIDGE adresse les objectifs suivants : au niveau politique, analyser, comparer et proposer une harmonisation des cadres institutionnels de politique de gouvernance des données ; au niveau intercommunautaire : développer des principes de pratiques de gestion et de partage à destination des chercheurs et les gestionnaires de données ; au niveau technique par la mise en œuvre de composants pour une meilleure interopérabilité des entrepôts.

Dans les objectifs principaux du projet ANR-BRIDGE, nous nous intéressons ici à la gouvernance des données initiée par les institutions de recherche présentes dans le projet. L'ambition du projet est de proposer un cadre de référence respectant l'environnement politique des données de recherche dans un cadre national, européen et international. Ce travail effectué dans le work package 1 (WP1) s'inscrit opportunément dans un contexte où de plus en plus d'éditeurs de revues scientifiques et d'agences de financement introduisent des politiques de données visant à augmenter la disponibilité des données de recherche à des fins de réutilisation.

II. Périmètre du work package 1 du projet ANR-BRIDGE

Nous introduisons au préalable quelques définitions pour positionner les travaux du work package 1 du projet ANR-BRIDGE.

1. Définitions

Politique des données

Une politique des données est un ensemble de principes généraux de haut niveau qui forment le cadre directeur dans lequel la gestion des données peut fonctionner (dans une organisation).

(traduit de «[OECD Glossary of Statistical Terms](https://doi.org/10.1787/9789264055087-en) » <https://doi.org/10.1787/9789264055087-en>, page 123)

Gouvernance des données

La gouvernance des données pose la question de la gouvernance de l'information (Koooper et al., 2011 ; Maurel, 2013) ; celle-ci fournit un cadre de responsabilité précisant les principes et règles de base, la structure de direction et les modalités de gestion de l'information, afin d'en assurer une utilisation efficace et efficiente.

Source : Da Sylva, Lyne, Dominique Maurel, Marc Bruyère, Marielle Saint-Germain, et Geneviève Gareau. 2019. « Écosystème de la gestion de données de recherche et professionnels de l'information: présentation des enjeux, de la méthodologie et des solutions préconisées d'une enquête canadienne ». Études de communication, langages, information, médiations, no 52 (juin): 51-70. <https://doi.org/10.4000/edc.8615>

« La gouvernance des données correspond à l'ensemble des organisations et des procédures mises en place au sein d'une entreprise [note des auteurs : ou d'un organisme] afin d'encadrer la collecte de données et leur utilisation. Il s'agit autant de respecter les obligations légales imposées par les pays et/ou l'Union européenne. »

Source : [Data Gouvernance ou gouvernance des données : Qu'est-ce que c'est ? \(lebigdata.fr\)](https://lebigdata.fr)

« Gouverner » les données signifie que des (bonnes) décisions sont prises à chaque étape du cycle de vie des données (de leur production ou réutilisation à leur valorisation) et que les responsabilités de chacun sont définies dans ce processus décisionnel.

Source : « Datapartage - Principes pour la gouvernance des données INRAE ». s. d. Consulté le 14 mai 2022. <https://datapartage.inrae.fr/Gouvernance-des-donnees2/Principes-pour-la-gouvernance-des-donnees-INRAE>.

La gouvernance des données n'est pas la gestion des données

« La gouvernance des données est souvent confondue avec d'autres termes et concepts relativement proches, dont la [gestion des données](#)¹».

Gestion des données de la recherche

La gouvernance des données intervient au niveau de la stratégie, des rôles, de l'organisation et des règles, alors que « la gestion des données de recherche - GDR (ou *Research Data Management - RDM en anglais*) est l'ensemble des processus [opérationnels] appliqués tout au long du cycle de vie d'un projet de recherche pour guider la collecte, la documentation, le stockage, le partage et la préservation des données de recherche. »

Source : [Alliance de recherche numérique du Canada](#)

2. Politique de science ouverte, politique des données de la recherche

Si des documents de **politique institutionnelle de science ouverte** commencent à être publiés, les documents spécifiques aux **politiques des données de la recherche** ne bénéficient pas encore d'analyses approfondies.

La *figure 1* montre comment une politique de science ouverte pour la recherche se positionne par rapport aux autres politiques d'ouverture des productions scientifiques, par exemple sur l'open access, l'éducation ouverte et les ressources éducatives libres.

Les travaux du projet ANR-BRIDGE étudient précisément les politiques d'ouverture et de gouvernance des données de la recherche, en relevant les grands principes présidant à leur mise en œuvre et en émettant des recommandations.

¹ <https://www.talend.com/fr/resources/guide-gouvernance-donnees/>

Relation entre politique de Science ouverte et politique des Données de la recherche - Périmètre Bridge

Picture 1. The policy in relation to other national open science documents.

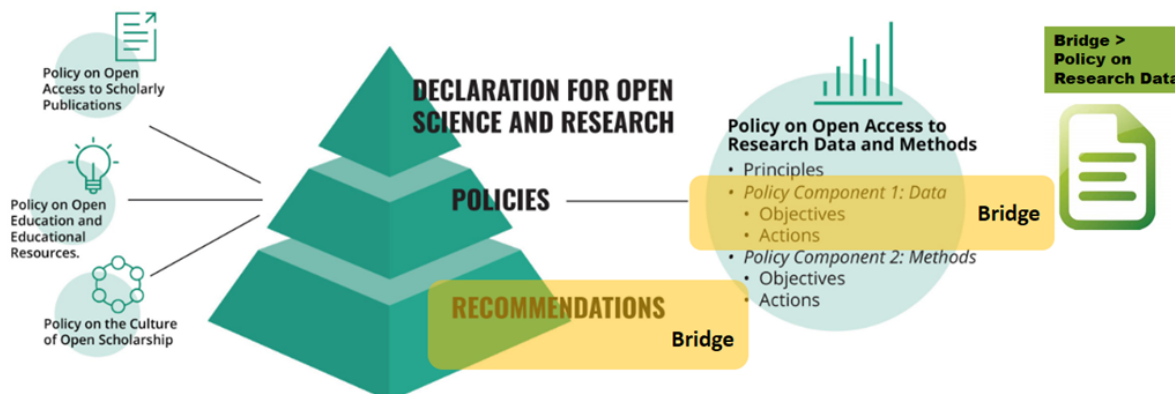


Figure 1 - Coordination, National Open Science and Research 2021. « Open Research Data and Methods National Policy and Executive Plan by the Higher Education and Research Community for 2021–2025 ». Responsible Research for Innovation. 2 201. <https://edition.fi/tsv/catalog/view/64/166/575-1>

III. Démarche du work package 1

Les travaux du WP1 du projet ANR-BRIDGE ont consisté dans un premier temps :

- à réunir les documents cadres et spécifiques qui formalisent des politiques ou des composantes de la gouvernance des données pour les trois établissements IRD, Cirad et INRAE. Ceci inclut également les documents mis à disposition des équipes de recherche (guides, recommandations techniques et organisationnelles, sites web intranet ou internet d'information), des documents utilisés par les établissements dans leurs procédures internes ou pour des cycles de formations; (cf. Tableau 1)
- à analyser le contexte politique national, européen et international de la gouvernance des données de la recherche (Plan national pour la science ouverte, guides du Comité pour la Science Ouverte, Horizon Europe, UNESCO, bailleurs internationaux ...)
- à établir une veille et à suivre des animations externes proposées tout au long du projet (conférences, webinaires nationaux ou internationaux RDA², CODATA³...) touchant au domaine.

² RDA - Research Data Alliance <https://www.rd-alliance.org/>

³ CODATA - Comité de données pour la science et la technologie <https://codata.org/>

Tableau 1- Exemple de documents cadres et spécifiques répertoriés qui formalisent des éléments de politique ou des composantes de la gouvernance des données des établissements

| Titre du document | Lien si disponible | Organisme |
|--|--|---|
| Enjeux éthiques et déontologiques de la gestion et du partage des données issues de la recherche | Avis numéro 8 sur les enjeux éthiques et déontologiques du partage et de la gestion des données issues de la recherche, Février 2016 https://www.inrae.fr/actualites/avis-partage-gestion-donnees-issues-recherche | Comité consultatif commun d'éthique INRAE-Cirad-Iframer-IRD |
| Feuille de route de la science ouverte - juin 2021 | https://www.ird.fr/une-feuille-de-route-pour-une-science-ouverte-et-partagee | IRD |
| Le Cirad est engagé dans le libre accès aux connaissances - 2021 | https://partage-connaissances.cirad.fr/textes-introduction/le-cirad-est-engage/le-cirad-est-engage-dans-le-libre-acces-aux-connaissances | CIRAD |
| Charte pour le libre accès aux publications et aux données | https://hal.inrae.fr/hal-02801732/document | INRAE |
| Principes de gouvernance des données pour les organismes de recherche - l'approche de l'INRAE | Aumont, Gilles, Michaël Chelle, Esther Dzale-Yeumo, Odile Hologne, Olivier Philipe, Hadi Quesneville, and Stéphanie Rennes. 2020. 'Principles of Data Governance for Research Organizations - INRAE's Approach'. In. https://hal.inrae.fr/hal-02955903 . | INRAE |
| La Gouvernance des données de la recherche au CIRAD : Document institutionnel | Document interne Février 2019 | CIRAD |
| La Gestion des données au CIRAD : Document de politique institutionnelle | Document interne Février 2019 | CIRAD |
| Organisation et mission des personnels responsables d'appui dans le cycle de vie des données | Lettres de mission selon organisation interne propre aux institutions (Documents internes) – Responsables science ouverte, – Administrateur fonctionnel et administrateur technique d'entrepôt de données Dataverse ... Politique d'emploi (GPEC) du domaine | IRD CIRAD INRAE |
| Propriété Intellectuelle, valorisation des données de la recherche.. | Chartes, documents internes | IRD CIRAD INRAE |
| Sites intranet et internet d'appui à l'intendance des données | https://data.ird.fr/ https://coop-ist.cirad.fr/gerer-des-donnees https://intranet-data.cirad.fr https://datapartage.inrae.fr/ | IRD CIRAD INRAE |
| Dispositifs de formation au domaine | Parcours de formations | IRD CIRAD INRAE |

Dans un deuxième temps, nous avons comparé les divers documents collectés et évalué leur alignement (convergences et divergences) afin d'identifier des critères et des règles communes tant en termes d'organisation, d'acteurs et de management d'une gouvernance des données explicite.

In fine, nous avons conclu collectivement que les différentes échelles d'étude, l'hétérogénéité des documents disponibles, les finalités et l'organisation inhérente à chaque établissement ne nous

permettaient pas d'obtenir une base pertinente pour une analyse comparative des politiques de gouvernance des données.

1. Quel cadre commun adopter ?

Quel cadre de référence pour appuyer une politique et gouvernance harmonisée des données de la recherche ?

Pour répondre à cette question, nous avons décidé de travailler à l'échelle d'une politique des données de la recherche, ce qui offre l'intérêt de fournir de grands principes directeurs pour fonder notre analyse. Plusieurs documents décrivent de manière opérationnelle les différents critères d'une politique de données de recherche comme i) le rapport FAIR-enabling Data Policy Checklist 2022. (Davidson et al. 2022)⁴, ii) des modèles de politiques de Science Ouverte et iii) des cadres de recommandations publiés dans le contexte des projets OpenAIRE ou EOSC-Pillar⁵ financés par l'Union Européenne.

Mais, comme le souligne le groupe d'intérêt de la Research Data Alliance « *Data Policy standardisation and implementation*⁶ » (RDA Data Policy), à mesure que la prévalence des politiques de données augmente, les chercheurs et le personnel de soutien des établissements de recherche risquent d'être déroutés par la communication d'exigences multiples, parfois même contradictoires.

C'est ainsi qu'en 2020, le groupe RDA Data Policy a élaboré une grille de lecture pointant quatorze critères caractérisés, définissant un cadre pour une politique des données de recherche des éditeurs des revues scientifiques, avec une étude actuellement en cours sur les financeurs de la recherche.

A noter qu'en 2021, le Collège des données du Comité pour la science ouverte⁷ a diffusé auprès des revues scientifiques des recommandations qui sont basées sur ces travaux⁸.

Contributions du projet ANR-BRIDGE

Le projet ANR-BRIDGE est parti de ce postulat : les caractéristiques d'une politique de données de la recherche des éditeurs scientifiques et des bailleurs de la recherche doivent être intelligibles par de multiples parties prenantes ; Ces caractéristiques peuvent alors contribuer à définir les termes d'une politique de données pour les établissements de recherche.

C'est pourquoi nous avons choisi d'adopter le cadre de référence du groupe RDA Data Policy, afin de converger vers des critères communs.

Ce cadre de référence nous permet également d'appréhender la complexité des différents mandats et le contexte national et international d'exercice des trois établissements de recherche.

Cadre de référence du groupe d'intérêt RDA Data Policy

⁴ Davidson, Joy, Grootveld, Marjan, Verburg, Maaïke, van Horik, René, O'Connor, Ryan, Engelhardt, Claudia, Garbuglia, Federica, Vieira, André, Newbold, Elizabeth, Proudman, Vanessa, & Horton, Laurence. (2022). FAIR-enabling Data Policy Checklist (1.0). Zenodo. <https://doi.org/10.5281/zenodo.6225775>

⁵ <https://www.eosc-pillar.eu/>

⁶ <https://www.rd-alliance.org/groups/data-policy-standardisation-and-implementation-ig>

⁷ Le Comité pour la science ouverte assure la mise en œuvre de la politique nationale:

<https://www.ouvrirlascience.fr/accueil/>

⁸ <https://www.ouvrirlascience.fr/recommandations-aux-revues-souhaitant-definir-une-politique-de-donnees-liees-aux-publications/>

Dans ses derniers travaux (nov. 2021), le groupe d'intérêt RDA Data Policy a travaillé sur l'*Élaboration d'un cadre de politique des données de recherche pour toutes les revues et éditeurs scientifiques*, [Hrynaszkiewicz et al., 2020]⁹.

Par l'examen des pratiques d'éditeurs scientifiques publiant collectivement plus de 10 000 revues, le groupe d'intérêt a caractérisé six profils types de politiques de données. Cette étude a été complétée par une première analyse des politiques de données des bailleurs de fonds, en s'appuyant sur le registre mondial des organismes de subventions maintenu par Crossref ¹⁰.

A partir d'un cadre normalisé, faisant l'objet d'un consensus, l'étude relève quatorze critères caractérisés (obligatoires, recommandés ou sans objet) définissant un cadre pour la politique de données de recherche des revues scientifiques et des financeurs de la recherche (voir Figure 2).

⁹ Hrynaszkiewicz, I., Simons, N., Hussain, A., Grant, R. and Goudie, S., 2020. Developing a Research Data Policy Framework for All Journals and Publishers. *Data Science Journal*, 19(1), p.5. <http://doi.org/10.5334/dsj-2020-005>

¹⁰ Crossref Funder Registry <https://www.crossref.org/services/funder-registry/> (licence CC0)

14 journal research data policy features arranged as 6 policy types (tiers)

| | Policy 01 | Policy 02 | Policy 03 | Policy 04 | Policy 05 | Policy 06 |
|--|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|
| Definition of the research data | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Exceptions to policy | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> | <input checked="" type="radio"/> |
| Embargoes | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> | <input checked="" type="radio"/> | <input checked="" type="radio"/> |
| Supplementary materials | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> | <input checked="" type="radio"/> | <input checked="" type="radio"/> |
| Data repositories | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> | <input checked="" type="radio"/> | <input checked="" type="radio"/> |
| Data citation | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> | <input checked="" type="radio"/> |
| Data licensing | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Researcher/ author support | <input checked="" type="radio"/> | <input checked="" type="radio"/> | <input checked="" type="radio"/> | <input checked="" type="radio"/> | <input checked="" type="radio"/> | <input checked="" type="radio"/> |
| Data availability statements | | <input type="radio"/> | <input checked="" type="radio"/> | <input checked="" type="radio"/> | <input checked="" type="radio"/> | <input checked="" type="radio"/> |
| Data formats and standards | | | | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> |
| Mandatory data sharing (specific data types) | | | | <input checked="" type="radio"/> | <input checked="" type="radio"/> | <input checked="" type="radio"/> |
| Mandatory data sharing (all papers) | | | | <input type="radio"/> | <input checked="" type="radio"/> | <input checked="" type="radio"/> |
| Peer review of data | | | | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> |
| Data Management Plans (DMPs) | | | | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Provide information

The text for the policy feature will be included in the policy template but it is clear that the feature will not be enforced and checked as part of the publishing or peer review process

Provide information and action

The text of the policy feature is included and makes clear where applicable that the feature will be checked and enforced in the publishing or peer-review process

Figure 2 - Quatorze caractéristiques de la politique relative aux données de recherche, classées en six types de politiques.

<https://datascience.codata.org/article/10.5334/dsj-2020-005/dsj-19-1086-g1.png/>

2. Méthode d'analyse du cadre de référence *RDA Data Policy*

Les travaux du projet ANR-BRIDGE se sont appuyés sur la revue initiale des documents cadres et spécifiques (cf. Tableau 1) et sur la connaissance des pratiques de gestion des données menées depuis plusieurs années au Cirad, à l'INRAE et à l'IRD.

Un premier atelier, organisé en novembre 2021, a permis de dresser un état des lieux initial et très large sur les données de la recherche dans nos organismes, en commentant notamment les quatorze critères

du cadre produit par le groupe RDA Data Policy. Cet inventaire s'est poursuivi avec une revue intermédiaire pour dégager les points saillants.

Un deuxième atelier en mars 2022, a permis d'affiner les sujets et d'aboutir à la proposition d'un nouveau profil « type » de politique de données destiné à un établissement de recherche (voir Figure 3), adossé au cadre commun défini par le groupe RDA Data Policy.

| Caractéristiques bailleurs de fonds & éditeurs | | | | | | | | Caractéristiques Établissement |
|--|--|-----------|-----------|-----------|-----------|-----------|-----------|-----------------------------------|
| n° | Caractéristiques | Policy 01 | Policy 02 | Policy 03 | Policy 04 | Policy 05 | Policy 06 | Ex. Policy 07 |
| 1 | Definition of the research data | ○ | ○ | ○ | ○ | ○ | ○ | ● |
| 2 | Exceptions to policy | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| 3 | Embargoes | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| 4 | Supplementary materials | ○ | ○ | ○ | ● | ● | ● | ● |
| 5 | Data repositories | ○ | ○ | ○ | ● | ● | ● | ● |
| 6 | Data citation | ○ | ○ | ○ | ○ | ● | ● | ● |
| 7 | Data licensing | ○ | ○ | ○ | ○ | ○ | ○ | ● |
| 8 | Researcher/ author support | ● | ● | ● | ● | ● | ● | ● |
| 9 | Data availability statements | | ○ | ● | ● | ● | ● | ○ |
| 10 | Data formats and standards | | | | ○ | ○ | ● | ● |
| 11 | Mandatory data sharing (specific data types) | | | | ● | ● | ● | ○ |
| 12 | Mandatory data sharing (all papers) | | | | ○ | ● | ● | ○ |
| 13 | Peer review of data | | | | ○ | ○ | ● | ○ |
| 14 | Data Management Plans (DMPs) | | | | ○ | ○ | ○ | ● |

Légende :

- = Information require
- = Information & action require
- = Sans objet

○ **Fournir l'information (recommandé)**
Le texte de la caractéristique sera inclus dans le modèle de politique, mais la caractéristique ne sera pas obligatoirement appliquée et vérifiée

● **Fournir l'information et agir (obligatoire)**
Le texte de la caractéristique est inclus dans le modèle de politique et indique clairement que la caractéristique sera appliquée et vérifiée

Bridge - Profil
Etablissement de recherche

Bridge retient toutes les caractéristiques pour encadrer une politique des données conforme au cadre des politiques des bailleurs de fonds et des éditeurs.

Chaque organisme choisit pour chaque caractéristique, un niveau d'exigence (obligatoire, recommandé, sans objet) selon sa position sur le critère.

Bridge définit
+ 2 nouvelles caractéristiques

Légal - réglementaire

Intégrité scientifique

Figure 3 - Profil d'un Établissement de recherche

Dans une phase finale d'alignement des politiques, le projet ANR-BRIDGE a documenté les critères (ou caractéristiques) en examinant les trois échelles nationale, européenne et internationale, de politiques publiques des données de la recherche.

Parmi les principaux documents relatifs aux politiques des données qui ont été examinés, citons :

- Au niveau national :
 - Le deuxième *Plan national pour la science ouverte (PNSO2) en 2021*
 - Le *Plan d'action 2022 de l'Agence Nationale pour la Recherche (ANR)*
 - *Vers le déploiement d'une politique commune en faveur de la science ouverte*, du Réseau des agences de financement françaises en faveur de la science ouverte ;
- Au niveau européen :
 - *Practical Guide to the International Alignment of Research Data Management - Extended Edition* de Science Europe¹¹
 - Le *Programme Guide, 2022 Horizon Europe (HORIZON)* ;
- Au niveau international :
 - Les travaux de synthèse du groupe RDA Data Policy IG
 - Et plus largement les recommandations de l'UNESCO sur une science ouverte émises en 2021.

¹¹ *Science Europe* est une association européenne qui sert de plateforme pour élaborer des positions sur les questions de politique de recherche et adresser des messages politiques aux institutions européennes, aux chercheurs, aux gouvernements nationaux et au public. <http://www.scienceeurope.org/>

Ces documents sont détaillés dans le tableau suivant :

Tableau 2 - Liste de documents relatifs aux politiques de données

| Pays Région | Organisation | Titre du document | Lien |
|---------------|---|---|---|
| France | Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation - MESRI | Deuxième Plan national pour la science ouverte, 2021-2024 (PNSO2) | https://www.ouvrirelascience.fr/deuxieme-plan-national-pour-la-science-ouverte/ |
| France | Agence Nationale de financement pour la Recherche – ANR | Plan d'action 2022 | https://anr.fr/fileadmin/documents/2021/PA-ANR-2022-V1.1a.pdf |
| France | Réseau des agences de financement françaises en faveur de la science ouverte | Vers le déploiement d'une politique commune en faveur de la science ouverte | https://anr.fr/fileadmin/documents/2022/Vers le deploiement d'une politique commune en faveur de la science ouverte.pdf |
| Europe | Science Europe, association représentant les grandes organisations de financement et d'organisations de la recherche, Bruxelles | Practical Guide to the International Alignment of Research Data Management - Extended Edition. <i>Guide pratique pour l'alignement international de la gestion des données de recherche - Edition étendue</i> | https://doi.org/10.5281/zenodo.4915862 |
| Europe | Horizon Europe, programme européen de soutien à la recherche et à l'innovation (2021-2027). | Horizon Europe (HORIZON), Programme Guide, 2022 | https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf |
| International | UNESCO, Organisation des Nations unies pour l'éducation, la science et la culture. | Recommandation de l'UNESCO sur une science ouverte, 2021 | https://unesdoc.unesco.org/ark:/48223/pf0000379949_fre |
| International | Data policy standardisation and implementation IG | Developing a Research Data Policy Framework for All Journals and Publishers | http://doi.org/10.5334/dsj-2020-005 |

IV. Cadre commun pour une politique des données de la recherche

Par convention nous conservons ici la numérotation des caractéristiques telle que présentée dans le cadre de référence du groupe RDA.

En première partie, nous présentons pour chaque critère une **définition** et un **panorama de l'existant**. Pour chacun d'eux, des **recommandations** sont formulées, qu'un établissement de recherche s'appropriera en fonction de son orientation sur le sujet.

En deuxième partie, nous émettons des **propositions** pour compléter ce cadre commun, avec notamment **deux critères supplémentaires** adaptés aux établissements de recherche.

1. Définitions des types de données (portée de la politique)

Ce critère définit les types de données de recherche concernés par la politique de données. Il permet de définir la portée de la politique (son périmètre) et de fournir des informations aussi bien générales que spécifiques à une discipline.

Panorama

Plusieurs définitions des données de la recherche coexistent :

- Le deuxième Plan national pour la science ouverte et l'Agence Nationale pour la Recherche considèrent tous deux la définition de l'Organisation de Coopération et de Développement Economiques (OCDE) : « Les données de la recherche sont définies comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider les résultats de la recherche. »¹².
- Le Guide du programme Horizon Europe, adopte plus largement la terminologie de résultats de la recherche qui sont des « résultats générés par l'action auxquels il est possible d'accéder sous forme de publications scientifiques, de données ou d'autres résultats et processus techniques tels que des logiciels, des algorithmes, des protocoles et des carnets électroniques ».
- Le document *What is Research data ?*¹³ de Australian Research Data Commons indique que « Le terme de données de la recherche désigne les données sous forme de faits, d'observations, d'images, de résultats de programmes informatiques, d'enregistrements, de mesures ou d'expériences sur lesquels un argument, une théorie, un test ou une hypothèse, ou un autre produit sur lesquels la recherche est basée ». « Elles peuvent être brutes, nettoyées ou traitées, ... ».
- Le groupe d'intérêt *Data Policy standardisation and implementation* de la RDA a étudié les cadres des politiques des données des éditeurs de revues scientifiques ainsi que des financeurs de la recherche. L'étude révèle que ces politiques s'appliquent aux données de recherche qui seraient nécessaires pour vérifier les résultats des recherches rapportées dans les articles publiés dans la revue. Les données de recherche comprennent les données produites par les auteurs ("données primaires") et les données provenant d'autres sources qui sont analysées par les auteurs dans leur étude ("données secondaires"). Les données de recherche comprennent tout matériel factuel enregistré qui est utilisé pour produire les résultats sous forme numérique et non numérique. Il s'agit

¹² Organisation de Coopération et de Développement Economique (OCDE), *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics*, Paris, 2007, 28 p., <http://www.oecd.org/fr/science/sci-tech/38500823.pdf>

¹³ https://guides.library.unisa.edu.au/ld.php?content_id=50662363 (page 3)

notamment de données tabulaires, de codes, d'images, de sons, de documents, de vidéos, de cartes, de données brutes et/ou traitées.

On notera ici que les enjeux liés à la Science Ouverte (ouverture, diffusion, contexte juridique) conduisent à rapprocher la gestion des algorithmes et des codes sources de celle des données.

Recommandations BRIDGE

R1. La définition des données de la recherche de l'OCDE constitue un socle minimal de définition partagée.

La définition proposée par l'OCDE dans son rapport *Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics* de 2007 est largement partagée :

“les données de la recherche sont définies comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider les résultats de la recherche”.

Bien qu'une telle définition soit un préalable à toute politique de gestion des données, il serait souhaitable que les établissements de recherche identifient son périmètre d'application, en précisant les différents types de données produites.

L'INIST (l'Institut de l'Information Scientifique et Technique) du CNRS définit ainsi cinq types de données :

- Données d'observation
- Données expérimentales
- Données de simulation
- Données dérivées ou compilées
- Données de référence.

Les données produites dans le cadre d'un projet de recherche passent par différentes phases durant leur cycle de vie et peuvent donc aussi être distinguées selon leur niveau de traitement :

- Données brutes : non interprétées et qui n'ont subi aucun traitement.
- Données traitées : produites après calibration/étalonnage ou correction des données brutes.
- Données dérivées ou compilées : agrégées, calculées, réorganisées, qui donnent une vue spécifique des données.
- Données analysées : résultant d'une analyse qualitative, statistique ou mathématique des données.

R2. Définir plus précisément les données de la recherche pour une meilleure lisibilité par les collectifs de recherche.

On pourra s'appuyer sur les éléments essentiels qui composent les définitions des données de la recherche, au niveau européen et international, référencés par le site DoRANum¹⁴ (Données de la Recherche : Apprentissage Numérique), une plateforme de formation portée par l'INIST-CNRS et le Groupement d'Intérêt Scientifique (GIS) réseau des URFIST.

De plus, DoRANum a intégré l'écosystème Recherche Data Gouv¹⁵, inauguré le 8 juillet 2022, amplifiant la visibilité de ses informations et des ressources disponibles au plan national.

¹⁴ <https://doranum.fr/plan-gestion-donnees-dmp/definitions-des-donnees-de-la-recherche/>

¹⁵ <https://recherche.data.gouv.fr>

R3. Privilégier une politique spécifique pour les algorithmes et les codes source produits par l'établissement. A défaut, les inclure dans le périmètre d'application de la politique de gestion des données au sens large du terme.

Le PNSO2 promeut une ouverture, une valorisation et une réutilisation des algorithmes et des codes sources issus de la recherche.

Si l'institution de recherche produit des logiciels scientifiques originaux, des modèles ou encore des outils d'aide à la décision, elle peut privilégier la définition d'une politique spécifique pour les algorithmes et le code source. Cette politique pourra énoncer les orientations de l'établissement tant sur la politique de préservation, de partage que sur les licences ou conditions de réutilisation dans le cadre d'activités de recherche y compris en partenariat (privé/public/nord/sud) en fonction de ses missions et finalités.

2. Les types de données à l'origine d'exceptions

Ce critère définit quelles données ne doivent pas être diffusées en accès libre. Les options alternatives pour décrire la disponibilité de ces données pourront être précisées.

Panorama

Dans le PNSO2, « l'obligation d'ouverture des données de la recherche publique, posée par la loi pour une République numérique de 2016, doit désormais se traduire dans les pratiques scientifiques grâce à des infrastructures et des services d'accompagnement adaptés. Elle est limitée par les exceptions légitimes encadrées par la loi, par exemple en ce qui concerne le secret professionnel, les secrets industriels et commerciaux, les données personnelles ou les contenus protégés par le droit d'auteur. Dans ces cas, les pratiques de partage des données devront être favorisées à travers la définition de protocoles maîtrisés. »

A l'instar des préconisations du PNSO2, les membres du réseau des agences de financement françaises (ADEME, ANR, Anses, INCa, Inserm/ANRS) s'engagent à favoriser l'ouverture des données de la recherche dans le respect du paradigme "aussi ouvert que possible, aussi fermé que nécessaire", en accord avec les efforts européens (Science Europe, Horizon Europe) et internationaux sur le partage des données de recherche. Ce principe signifie que des exceptions à cette règle peuvent être faites dès lors qu'elles sont justifiées.

Pour ce critère, le groupe RDA Data Policy précise : « une telle politique n'exige pas le partage public de données quantitatives ou qualitatives qui pourraient permettre d'identifier un participant à la recherche, à moins que les participants n'aient consenti à la divulgation des données. Cette politique n'exige pas non plus le partage public d'autres données sensibles, telles que l'emplacement d'espèces menacées. »

Les alternatives au partage public de données sensibles ou personnelles comprennent :

- Le dépôt des données de recherche dans des entrepôts dont l'accès est contrôlé ;
- L'anonymisation des données avant leur partage public ;
- Le seul partage des métadonnées relatives aux données de recherche ;
- La mention des procédures d'accès aux données de recherche dans l'article qui leur est associé et la gestion des demandes d'accès aux données émise par les autres chercheurs, qui sont précisément l'objet du paragraphe « Data Availability Statement » inclus dans un article.

Recommandations BRIDGE

R4. Appliquer le principe « aussi ouvert que possible, aussi fermé que nécessaire » dans un cadre juridique et éthique adapté.

- ✓ Il s'agit de reconnaître que certaines données ne peuvent pas être ouvertes, sur la base d'exceptions légitimes pour des raisons contractuelles, de confidentialité, réglementaires, de sécurité publique, de restrictions éthiques, de droits de propriété intellectuelle ou d'intérêts commerciaux.

3. Données sous embargo

Le droit des chercheurs à la primeur de l'utilisation des données générées durant leurs recherches est un principe largement accepté du partage des données. Mais, les durées d'embargo considérées comme raisonnables peuvent varier selon la discipline, le type de données et de recherche.

Panorama

Pour ce critère, il s'agit de définir si des embargos sont permis pour la diffusion des données et lesquels. Un embargo peut s'appliquer lors de la publication, pour des raisons de protection de la propriété intellectuelle, de dépôt de brevet par exemple.

Ce type d'information n'est souvent pas mentionné dans les documents étudiés, car il est trop spécifique à une discipline ou une revue.

Dans le cas précis des données associées à une publication, la période d'embargo pourrait être celle de la publication. Ainsi, l'Union Européenne préconise de limiter l'embargo sur les publications à six mois en sciences, techniques et médecine et à douze mois en sciences humaines et sociales¹⁶.

Pour le projet européen FAIRsFAIR¹⁷, la question des périodes d'embargo autorisées doit être clairement traitée dans une politique de données.

Recommandation BRIDGE

R5. Les embargos sur le partage des données sont permis dans la limite de durées raisonnables, en accord avec les bonnes pratiques de la discipline scientifique dans laquelle s'inscrivent les données et le projet.

Un embargo peut s'appliquer en particulier aux données primaires.

La durée d'embargo doit être documentée.

Cette durée doit être établie suivant les bonnes pratiques des communautés scientifiques.

4. Documents supplémentaires (*supplementary materials*)

Certaines revues scientifiques continuent à promouvoir le partage des données sous la forme de documents supplémentaires (*supplementary materials*, annexes, ...) attachés aux articles.

¹⁶https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm

¹⁷ The Project | FAIRsFAIR <https://www.fairsfair.eu/the-project>

Ce critère définit le positionnement d'une politique de données d'établissement de recherche vis-à-vis d'une telle pratique.

Panorama

De nombreuses politiques de données (bailleurs de la recherche) indiquent que les données doivent être déposées dans un entrepôt de données de confiance (cf. §7. *Entrepôts de données*). A l'opposé et au niveau des revues scientifiques, la recommandation de déposer les données sous forme de « documents supplémentaires » (exemple : tableaux de données au format PDF) est une pratique encore assez répandue. Ces pratiques ne permettent pas la FAIRisation¹⁸ des données et en particulier leur « accessibilité ». En pareil cas, les données peuvent aussi avoir le même statut que l'article scientifique pour lequel les auteurs peuvent avoir cédé leurs droits à l'éditeur.

Recommandation BRIDGE

R6. La publication de données en tant que documents supplémentaires (Supplementary Materials) est déconseillée. Voir aussi §7. *Entrepôts de données*.

- ✓ Les données de recherche doivent être déposées dans des entrepôts qui respectent les critères de confiance. Ces entrepôts offrent une meilleure découverte des données (facilité de recherche) et augmentent leur capacité à être réutilisées et citées¹⁹.

5. Entrepôts de données

Le dépôt des données dans un entrepôt de données de la recherche est à privilégier pour favoriser leur partage.

Lorsqu'ils sont disponibles et reconnus de confiance, les entrepôts de données spécifiques à une communauté scientifique ou à une discipline sont préférés aux entrepôts généralistes.

Panorama

Le *Plan national pour la science ouverte* (PNSO2) encourage la structuration et l'ouverture des données de la recherche en recommandant leur diffusion suivant les principes FAIR, dans des entrepôts de données certifiés. Sous l'impulsion du PNSO2, l'organisation de l'écosystème français des infrastructures de la science ouverte, s'articule avec l'écosystème européen et international. Les notions de certification et de confiance des entrepôts sont abordées dans les recommandations.

Pour le dépôt des données de recherche, l'ANR invite les coordinateurs ou coordinatrices de projets à consulter la politique de science ouverte de leur établissement, leur cellule d'appui aux montages des projets, le portail DMP OPIDoR (ressources Plan de Gestion de Données) ainsi que des ressources disciplinaires comme celles de la Research Data Alliance²⁰.

Science Europe recommande le dépôt des données dans des **entrepôts de confiance** où les chercheurs peuvent les stocker à des fins de partage.

Horizon Europe, recommande que : dans la mesure où les données sont autorisées à être partagées, les projets de recherche financés par ce programme les déposent dans un référentiel possédant les caractéristiques des dépôts de confiance, et le justifient en conséquence dans leur plan de gestion des données.

¹⁸ ANR BRIDGE (2023) Principaux concepts techniques. 8 p. Zenodo. <https://doi.org/10.5281/zenodo.7474725>

¹⁹ Piwowar and Vision (2013), Data reuse and the open data citation advantage. PeerJ 1:e175; DOI [10.7717/peerj.175](https://doi.org/10.7717/peerj.175)

²⁰ <https://www.rd-alliance.org>

Une liste de critères de choix d'entrepôts de confiance sont relayés par exemple, par des organisations comme l'IRD²¹.

Recommandations BRIDGE

R7. Mettre en place une politique des données de recherche incite un établissement à identifier des entrepôts de données de confiance au sein de l'écosystème national, européen ou international.

Les communautés scientifiques doivent s'interroger sur le choix d'entrepôts disciplinaires en fonction des objectifs des projets et de la nature des partenariats (Nord/Sud/privé/public).

Pour établir et documenter ce choix, l'élaboration systématique de Plans de gestion des données pour les projets de recherche offre un espace d'échanges et d'accords entre les parties prenantes en amenant à étudier les différents choix d'entrepôts de données qui s'offrent au projet.

R8. Communiquer sur les critères de sélection d'un entrepôt de données de confiance auprès des scientifiques.

Il existe plusieurs critères pour déterminer le niveau de confiance d'un entrepôt de données ; les critères de certification du CoreTrustSeal²² ou les principes TRUST²³ par exemple.

Science Europe dans son *Guide pratique pour une harmonisation de la gestion des données de recherche*²⁴, présente les critères minimaux structurés autour de quatre piliers majeurs que tous les entrepôts de confiance doivent satisfaire :

- 1) Fourniture d'identifiants uniques et pérennes ;
- 2) Des métadonnées descriptives, normées fournissant une information pertinente, lisibles par l'homme et la machine ;
- 3) Un accès aux données et aux licences d'utilisation ;
- 4) La préservation pour assurer la persistance des métadonnées et des données dans le temps.

La communication doit privilégier les critères de la certification CoreTrustSeal qui s'impose progressivement dans le paysage. Ce point est développé dans la recommandation qui suit.

R9. Soutenir le processus de certification CoreTrustSeal des entrepôts de données français.

Dans le cadre des recommandations du PNSO, les entrepôts de données doivent s'engager dans une certification CoreTrustSeal²⁵. Celle-ci porte sur les enjeux de qualité des données, des processus de gestion et de préservation d'un entrepôt. Elle atteste de la fiabilité des services de l'entrepôt de données où les scientifiques peuvent déposer et réutiliser des données scientifiques en confiance, à court, moyen et long terme.

Le référentiel CoreTrustSeal (CTS : 2020-2022) est fondé sur seize critères d'évaluation relatifs aux dépôts de données dignes de confiance. Il a été révisé dans sa nouvelle version (2023-2025), en

²¹ <https://data.ird.fr/entrepots-de-donnees-2/>

²² « Entrepôts de données de confiance : critères de conformité ». Consulté le 20 janvier 2023.

<https://www.ouvri.lascience.fr/entrepots-de-donnees-de-confiance-criteres-de-conformite>.

²³ Lin, Dawei, Jonathan Crabtree, Ingrid Dillo, Robert R. Downs, Rorie Edmunds, David Giarretta, Marisa De Giusti, et al. « The TRUST Principles for Digital Repositories ». *Scientific Data* 7, n° 1 (14 mai 2020): 144. <https://doi.org/10.1038/s41597-020-0486-7>.

²⁴ « Guide pratique pour une harmonisation internationale de la gestion des données de recherche ». Consulté le 31 mars 2022. <https://www.ouvri.lascience.fr/guide-pratique-pour-une-harmonisation-internationale-de-la-gestion-des-donnees-de-recherche> et en version originale anglaise :

SCIENCE EUROPE. « Practical Guide to the international alignment of research Data Management », 2018. https://www.scienceurope.org/media/jezkhnoo/se_rdm_practical_guide_final.pdf.

²⁵ <https://www.coretrustseal.org/>

renforçant ces critères d'évaluation. A ce jour trois entrepôts français de données scientifiques sont certifiés CTS, comme présenté sur le site Internet de l'organisation CoreTrustSeal²⁶.

R10. Fournir un référentiel ou liste des entrepôts de données de confiance recommandés.

Afin d'aider les collectifs scientifiques dans leur choix de dépôts de données, il s'agirait d'établir une sélection d'entrepôts de confiance en s'appuyant sur les critères mentionnés précédemment.

Pour réaliser cet objectif, il est essentiel d'associer les scientifiques pour les sensibiliser sur ces critères et dresser une liste d'entrepôts adaptés aux besoins des communautés.

R11. En l'absence d'entrepôt de confiance disciplinaire ou thématique, déposer dans un entrepôt de confiance pluridisciplinaire national, international ou institutionnel.

Plusieurs établissements de recherche (Cirad, INRAE, Ifsttar, IRD, Sciences PO, Université de Lorraine...), ont mis en place un entrepôt institutionnel de données de la recherche. Ce service répond au besoin de nombreux scientifiques qui mènent leurs recherches avec des données dites de longue traîne ou sans entrepôt adapté. Pour le Cirad et l'IRD, les entrepôts ont aussi pour objectif de valoriser et préserver les données coproduites avec des partenaires du Sud.

Dans la mise en œuvre du PNSO2 et pour favoriser les pratiques actives d'ouverture des données, l'**entrepôt Recherche Data Gouv**²⁷ offre depuis l'été 2022, un entrepôt de données pluridisciplinaire qui vient compléter les infrastructures nationales et européennes déjà utilisées par certaines communautés disciplinaires. Les Centres de Référence Thématiques²⁸ sont une composante de l'écosystème Recherche Data Gouv pour accompagner ces communautés scientifiques. Enfin, un catalogue national signalera bientôt les données déposées sur d'autres entrepôts de confiance, constituant ainsi à terme le point de visibilité de l'ensemble des données de la recherche française.

6. Citation des données

Ce critère indique si le principe de la citation des données est encouragé ou rendu obligatoire par l'établissement de recherche.

La politique de données encourage ou rend obligatoire i) l'attribution d'identifiants permanents pour retrouver les données de recherche et ii) soutient la citation de données en tant que référence formelle.

Panorama

La citation des données est l'un des éléments fondateurs des principes FAIR et des principes « Data Citation »²⁹. Pour pouvoir citer un jeu de données, il faut disposer au minimum d'un identifiant permanent (tel qu'un DOI) et des informations suivantes :

Auteurs (créateurs), Date, Titre, Editeur (entrepôt), Doi ou identifiant. Version.

Sur le sujet, voir également la fiche CoopIST³⁰.

²⁶ <https://www.coretrustseal.org/why-certification/certified-repositories/>

²⁷ Recherche Data Gouv, un écosystème national au service du partage et de l'ouverture des données de recherche <https://recherche.data.gouv.fr/>

²⁸ <https://recherche.data.gouv.fr/fr/page/centres-de-reference-thematiques-expertises-par-domaine-scientifique>

²⁹ Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014 <https://doi.org/10.25490/a97f-egykh>

³⁰ Deboin, M.C. 2021. Citer un jeu de données scientifiques, en 4 points. Montpellier (FRA) : CIRAD, 4 <https://doi.org/10.18167/coopist/0058>

Recommandation BRIDGE

R12. Citation de données

- ✓ Définir les auteurs du jeu de données en identifiant tous les rôles possibles, toutes les contributions (conception : idées et outils, collecte, curation, traitement, contrôle/vérification, supervision, ...). Les auteurs d'un jeu de données peuvent être différents de ceux d'un article associé³¹.
- ✓ Veiller à citer le nom de l'entrepôt où sont déposés les données.
- ✓ Dans les métadonnées, il est recommandé de renseigner, les organismes et les structures (unités de recherche) d'affiliation des auteurs selon les règles en vigueur, souvent définies dans les conventions d'UMR, afin de faciliter leur repérage, produire des listes et des indicateurs (institutionnels, etc.). De même, il est recommandé de renseigner l'identifiant ORCID de chaque auteur.

7. Licences des données

Ce critère définit les règles de l'établissement sur les licences et les droits d'auteur s'appliquant aux données de recherche.

Panorama

Une licence (ou plus généralement, les conditions de réutilisation des données) qu'elle soit ouverte ou fermée, doit être choisie en amont de la production des données entre les parties prenantes dans le cadre d'un accord de consortium, d'un cadre contractuel, d'un Plan de Gestion de Données (PGD).

Dans un même jeu de données et selon les conditions générales d'utilisation de l'entrepôt, il peut arriver que différentes licences soient attribuées en fonction du type de contenus (par type de fichier ou par groupes de fichiers).

Dans le cas de données ouvertes, nous avons identifié des pratiques différentes selon les organismes. L'INRAE privilégie par défaut la licence Etalab, l'IRD et le Cirad qui évoluent plus spécifiquement dans un cadre international ont choisi d'attribuer par défaut les licences Creative Commons³².

Les revues encouragent la mise à disposition des données de recherche sous des licences ouvertes qui permettent une libre réutilisation.

Le cas des données non ouvertes (par accord ou cadre contractuel) entre dans le cas des exceptions. §2. *Les types de données à l'origine d'exceptions.*

³¹ Deboin, M.C. 2022. Reconnaître tous les contributeurs d'une publication. Montpellier (FRA) : CIRAD, 5 p. <https://doi.org/10.18167/coopist/0007>

³² Fily, M.F. 2015. Connaître et utiliser les licences Creative Commons. Montpellier (FRA) : CIRAD, 11 p. <https://doi.org/10.18167/xtnv-d457>
<https://creativecommons.org/licenses/?lang=fr-FR>

Recommandation BRIDGE

R13. Licences des données

- ✓ Il est conseillé d'utiliser des licences reconnues qui offrent l'avantage d'être établies juridiquement, interprétables par des machines et répondant aux principes FAIR. Les licences les plus utilisées sont la Licence Ouverte Etalab (France) et les licences Creative Commons (internationales). Des compatibilités peuvent exister entre certaines licences comme entre la Licence Ouverte Etalab et les standards internationaux (ODC-BY, CC-BY 2.0.³³).
- ✓ Le choix de l'entrepôt de données peut restreindre le choix de licences. Vérifier les conditions d'attribution de licence en amont du dépôt.
- ✓ La licence attachée à chaque jeu de données est discutée à l'avance et mentionnée dans le Plan de Gestion de Données du projet.
- ✓ Les coproducteurs ou responsables des données doivent s'accorder sur le type de licence d'utilisation de chaque jeu de données. Attention aux règles pouvant varier d'un partenaire à l'autre en s'appliquant à un jeu de données produit en commun.

Si le choix s'oriente vers une licence CC0, porter attention aux conséquences telle que l'abandon de tous les droits et en particulier du droit d'être cité.

Depuis la diffusion des recommandations au niveau national comme celles du « Guide d'analyse du cadre juridique en France, 2017 » dans le contexte de l'ESR, la mise en application de la Loi pour une République numérique pour les données de la recherche fait l'objet de publications³⁴.

8. Soutien aux chercheurs | auteurs

L'ouverture et le partage des données de la recherche est une nouvelle approche.

Ce critère définit les dispositions adoptées par l'établissement sur l'accompagnement des scientifiques dans cette nouvelle démarche : des informations sont disponibles, des personnes contacts sont identifiées pour obtenir un appui tel que sur le respect des politiques en la matière, la recherche de dépôts, les dispositions légales, les licences, les droits de propriété intellectuelle ou pour répondre aux questions courantes.

Panorama

L'INRAE, le Cirad et l'IRD ont développé depuis plusieurs années des dispositifs d'accompagnement pour leurs scientifiques sur la gestion et la diffusion des données : documents et sites web, formations en ligne et à distance, séminaires. Depuis quelques années de nombreuses universités et organismes proposent ce type de service et les scientifiques des UMR ont souvent le choix entre les services de plusieurs établissements.

Le projet national *Recherche Data Gouv*³⁵, inauguré le 8 juillet 2022 a pour ambition de fédérer les dispositifs d'accompagnement des équipes de recherche sur tout le territoire en créant les ateliers de la donnée. Sa mise en œuvre est inscrite dans le deuxième Plan national pour la science ouverte avec

³³ <https://www.etalab.gouv.fr/licence-ouverte-open-licence/>

³⁴ Cécile Arènes, Lionel Maurel, Stephanie Rennes. Guide d'application de la Loi pour une République numérique pour les données de la recherche. Comité pour la science ouverte. 2022. ([hal-03968218](https://hal.archives-ouvertes.fr/hal-03968218))

³⁵ Recherche Data Gouv, un écosystème au service du partage et de l'ouverture des données de recherche <https://recherche.data.gouv.fr/fr/page/a-propos-de-recherche-data-gouv>

la Politique des données, des algorithmes et des codes sources du ministère de l'Enseignement supérieur et de la Recherche.

« Les **ateliers de la donnée**³⁶ en région ont pour vocation d'être le point d'entrée de proximité à destination des équipes de recherche pour toute question relative à la gestion et à la diffusion des données de la recherche. Pour apporter une aide aux équipes de recherche, les ateliers proposent des services généralistes d'appui aux problématiques de gestion, structuration, partage et ouverture des données de recherche. Ces ateliers n'ont pas vocation à se substituer aux services d'accompagnement (locaux, nationaux ou internationaux) déjà existants dans certains établissements ou dans certains domaines disciplinaires et thématiques. Les ateliers de la donnée pourront s'appuyer sur les dispositifs disciplinaires et/ou thématiques identifiés comme **centres de référence thématiques**³⁷ de *Recherche Data Gov.* »

Recommandation BRIDGE

R14. Les dispositifs d'accompagnement généralistes d'appui sur la gestion, la structuration et la diffusion des données devraient être associés ou intégrés aux ateliers de la donnée de Recherche Data Gov.

- ✓ Échanger sur les pratiques entre universités et établissements de recherche. Harmoniser, mutualiser les approches généralistes quand cela est pertinent. Par leur spécialisation, certains dispositifs d'accompagnement institutionnels existants sont complémentaires aux ateliers de la donnée (exemples : partenariats international, données de santé humaine).
- ✓ Conserver des statistiques des activités d'accompagnement pour un suivi dans le temps.

9. Déclarations de disponibilité des données (*Data Availability Statement*)

Ce critère définit la règle donnée par l'établissement sur la fourniture de déclarations de disponibilité des données (DAS).

Panorama

Une *déclaration de disponibilité des données (Data Availability Statement - DAS)* est une manière simple et cohérente d'exprimer la disponibilité des données ouvertes - ou dans certains cas de justifier de leur indisponibilité pour des raisons légales ou éthiques.

Par défaut le DAS contient la citation du jeu de données et les conditions de disponibilité (contacts, modalités, licence). Il doit être bien rédigé et contenir la citation complète, standardisée avec le DOI, pour faciliter la découverte des données.

Les revues scientifiques exigent de plus en plus souvent un texte rédigé de déclaration d'accessibilité aux données ouvertes, inséré dans l'article. Certains organismes de financement encouragent dans leur politique de données cette pratique, par exemple la Fondation Bill et Melinda Gates (US).

³⁶ <https://recherche.data.gouv.fr/fr/page/ateliers-de-la-donnee-des-services-generalistes-sur-tout-le-territoire>

³⁷ <https://recherche.data.gouv.fr/fr/page/centres-de-reference-thematiques-expertises-par-domaine-scientifique>

Recommandation BRIDGE

R15. Déclaration d'accessibilité aux données (Data Availability Statement - DAS)

- ✓ Le DAS est une bonne pratique qui devrait accompagner toutes les publications qui s'appuient sur des données.
- ✓ Mettre à disposition des modèles de DAS, informer sur son usage auprès des scientifiques.
- ✓ Le DAS ne remplace pas la saisie de métadonnées dans l'entrepôt où les données sont publiées. En effet, les métadonnées permettent de disposer des mêmes informations sous un format plus FAIR (notamment la lisibilité par les machines).
- ✓ La génération de DAS pourrait être automatisée en s'appuyant sur les métadonnées.
- ✓ Le DAS en lui-même ne constitue pas une publication scientifique. Pour ce faire, s'orienter vers la rédaction d'un Data paper.

10. Formats et normes de (méta)données

Ce critère indique le choix de l'établissement sur la recommandation de normes de données et de métadonnées spécifiques aux communautés thématiques et disciplinaires, si leur usage est encouragé, requis dans certains cas, ou dans tous les cas.

Le critère indique également si certains formats de fichiers, tels que les formats ouverts, sont recommandés ou requis.

Panorama

Les données de recherche préparées selon des formats ouverts³⁸ et des normes de données communautaires sont plus réutilisables, interopérables et accessibles.

Dans le paysage national, les *centres de référence thématiques* développent une expertise dans la gestion, la structuration et la diffusion de données pour leur communauté scientifique et participent à la définition des normes internationales de description et de diffusion des données du domaine.

En général, les éditeurs de revues ne mentionnent pas explicitement les principes FAIR dans leur politique de données. Certaines revues encouragent les auteurs à partager les données de leur recherche en utilisant des formats et des normes de données et de métadonnées reconnus par leur communauté de recherche en les invitant à consulter le site de référence FAIRsharing.org³⁹ pour plus d'informations sur les formats et les normes de partage de données établis.

Recommandation BRIDGE

R16. Une politique de données encourage les auteurs à partager les données scientifiques en utilisant des formats ouverts et des normes de données et de métadonnées FAIR reconnus par les communautés scientifiques.

- ✓ Veiller et suivre les recommandations et actions menées par les *centres nationaux de référence thématiques* et les communautés scientifiques internationales y compris en cours de structuration.

³⁸ [Format ouvert ou fermé ? – DoRANum](#)

³⁹ <https://fairsharing.org/>

Quelques référentiels d'appui relevés par le projet ANR-BRIDGE pour les communautés scientifiques :

- Consulter le site FAIRsharing.org pour les informations sur les formats et les normes de partage de données établis.
- Effectuer une veille sur les travaux communautaires portés au sein de la Research Data Alliance, RDA for Disciplines⁴⁰.
- Pour anticiper et appuyer un processus d'archivage à long terme se référer aux recommandations de formats ouverts du CINES⁴¹ - Centre Informatique National de l'Enseignement Supérieur (mission archivage pérenne documents et données numériques).

11. Partage obligatoire de types de données spécifiques

Ce critère indique si le partage des données est obligatoire pour des types spécifiques de données de recherche.

Panorama

Lorsqu'il existe des mandats communautaires établis pour le partage des données et des mécanismes par lesquels ces types de données doivent être partagés, la politique de données peut préciser le respect de normes et de cadres communautaires.

Recommandation BRIDGE

R17. Partage obligatoire de types de données spécifiques

- ✓ Si une pratique d'ouverture communautaire thématique et disciplinaire est bien établie et supportée, il est pertinent de le préciser dans une politique de données.
- ✓ Ce critère est à mettre en perspective avec les actions des centres de référence thématiques à portée nationale et disciplinaire portés par des communautés scientifiques qui structurent leurs dispositifs pour soutenir la gestion, le traitement, le partage et l'ouverture des données de leur domaine thématique et scientifique.

12. Partage obligatoire de données

Ce critère définit la règle de l'établissement dans le cas où le partage des données est une condition de soumission pour un projet ou un article scientifique.

Panorama

Dans le cas des projets ANR ou Européens (Horizon Europe, bourses ERC...) les bénéficiaires s'engagent à garantir le libre accès immédiat aux publications scientifiques et à adopter une démarche d'ouverture des données de recherche conforme au principe « aussi ouvert que possible, aussi fermé que nécessaire ». Certains bailleurs comme la Fondation Gates exigent une ouverture de toutes les données produites, sauf exception dont les raisons sont alors documentées.

⁴⁰ [RDA for Disciplines | RDA \(rd-alliance.org\)](https://www.fairsharing.org/)

⁴¹ <https://www.cines.fr/archivage/des-expertises/les-formats-de-fichier/>

De plus en plus de revues exigent également le partage et l'ouverture des données à la fois pour les relecteurs de l'article, mais également pour les lecteurs, pour favoriser leur réutilisation.

Certains éditeurs scientifiques proposent leurs propres entrepôts de données pour partager les données.

Les revues scientifiques et les agences de financement peuvent mentionner les types de données qui rentrent dans cette obligation.

Recommandation BRIDGE

R18. Partage obligatoire de données

- ✓ Vérifier en amont l'exigence de partage des données par la revue scientifique, le financeur du projet, un accord de projet ou de consortium.
- ✓ Pour le partage des données se référer :
 - au critère §4 - Types de données à l'origine d'exceptions
 - au critère §5 - Entrepôts de données
 - et au critère §6 - Documents supplémentaires.

13. Examen des données par les pairs

Ce critère indique si l'examen des données par des pairs est prévu ou requis dans certaines conditions de soumission et de diffusion des données.

Panorama

Lorsque les données sont mises à disposition avec les articles de recherche, elles sont accessibles pour une évaluation par les pairs.

Pour les revues qui mettent fortement l'accent sur les données, comme les Data papers, un examen cohérent des données pour la validation et la réutilisation peut améliorer la reproductibilité et la qualité des recherches publiées. Les instructions ou les grilles d'évaluation⁴² destinées aux examinateurs précisent parfois explicitement de noter la qualité du jeu de données associé à un article.

Recommandation BRIDGE

R19. Le projet ANR-BRIDGE relève qu'un examen par les pairs des données destinées à être ouvertes pourrait être initié au sein des collectifs de recherche

- ✓ L'approche pourrait être adoptée dans le cadre d'une révision avant la soumission de Data papers.

⁴² Par exemple : « Guidelines for Reviewers: How to Review a Data Paper? », 23 août 2019. <https://ist.blogs.inrae.fr/afs/2019/08/23/guidelines-for-reviewers-how-to-review-a-data-paper/>.

14. Plan de Gestion de Données

Ce critère définit les règles de l'établissement sur l'élaboration des Plans de Gestion de Données (PGD).

Panorama

La planification de la gestion des données est une première étape essentielle du cycle de vie d'un projet de recherche. Très vite il est apparu que le meilleur niveau où l'on pouvait i) limiter les risques éthiques, juridiques et de préservation, ii) formaliser les aspects de provenance, de production, de co-production en partenariats, de traçabilité des données (producteur(s), contributeur(s), auteurs(s)...) et iii) de « FAIRisation ou de FAIRification⁴³ » étaient les plans de gestion de données.

Dans un projet, le plan de gestion de données (ou *Data Management Plan - DMP*) est « l'outil » le plus adapté pour la gestion, le suivi et le contrôle des données par les partenaires.

- Le PGD n'est pas exigé par les éditeurs.
- Le PGD est exigé par les organismes de financement français et européens (réseau des agences de financement françaises, programme Horizon Europe, ...).

Les trois partenaires du projet ANR-BRIDGE, INRAE, Cirad et IRD recommandent l'élaboration d'un PGD dans le cadre des projets et l'utilisation de la plateforme DMP OPIDoR⁴⁴ avec des instructions fournies en appui.

Il est conseillé de prévoir des mises à jour du PGD tout au long du cycle de vie de la recherche.

A noter que l'INRAE a initié un modèle de Plan de Gestion de Données de structure (unité de recherche, plateau technique, infrastructure...) destiné à documenter son activité en matière de données en mutualisant l'ensemble des informations utiles pour les utilisateurs de cette structure. Ce PGD peut être mis à disposition par la structure sur son site web⁴⁵ ou sur un espace dédié sur un entrepôt de données⁴⁶ par exemple.

Dans un projet spécifique, ces informations peuvent être alors exploitées et reprises dans le PGD.

Le PGD ne s'applique pas qu'aux projets où les données sont ouvertes : « Un PGD peut être établi aussi bien dans une optique de partage des données que pour des données en accès restreint ou fermé, total ou partiel. Le PGD mentionnera dans ce cas les raisons de non partage »⁴⁷.

Recommandations Bridge

R20. Usages des Plans de gestion de données

- ✓ Généraliser le Plan de Gestion de Données pour les projets, mais aussi pour les thèses.
- ✓ La politique de données doit indiquer clairement que le PGD doit être mis à jour au cours de la durée de vie du projet. Les PGD mis à jour doivent être clairement identifiés par leur version.
- ✓ Pour élaborer un PGD, s'appuyer sur les documents internes ou externes signalés dans les recommandations de votre organisation et définissant le cadre législatif, juridique et éthique de vos données.

⁴³ Processus de FAIRification adopté par GO FAIR <https://www.go-fair.org/fair-principles/fairification-process/>

⁴⁴ Portail et outil national de rédaction de PGD porté par l'INIST-CNRS) <https://dmp.opidor.fr/>

⁴⁵ Exemple URGI : <https://urgi.versailles.inrae.fr/Platform/General-Terms-of-Use/Conditions-Generales-d-Utilisation>

⁴⁶ Exemple URGI : Michotey, Celia, 2021, "Plant Bioinformatics Facility data management plan", <https://doi.org/10.15454/9HM5UI>, Recherche Data Gouv, V1

⁴⁷ https://doranum.fr/plan-gestion-donnees-dmp/plan-de-gestion-des-donnees-fiche-synthetique_10_13143_cg4-0k53/

- ✓ Offrir un service de relecture du PGD dans la mesure du possible.
- ✓ Utiliser s'il est disponible, le modèle de PGD du bailleur de fonds.
- ✓ Conseiller un modèle de PGD par défaut à utiliser.
- ✓ Consulter les instructions d'appui dans DMP OPIDoR.
- ✓ Déposer les PGD dans une archive ouverte afin de pouvoir y accéder de façon pérenne.
- ✓ Considérer l'utilisation de modèles « DMP machine actionnable » de DMP OPIDoR

Les politiques de données des établissements de recherche devraient encourager les chercheurs et le personnel d'appui à identifier collectivement les coûts pertinents de Gestion des Données de la Recherche FAIR (GDR FAIR). Ces coûts devraient être demandés dans les demandes de subvention, par l'élaboration d'un plan de gestion des données élaboré en amont.

R21. La politique indique clairement que les coûts justifiés associés à une Gestion des Données de la Recherche FAIR doivent être identifiés par la préparation d'un PGD, préalable à la soumission du projet.

V. Autres éléments pour appuyer une politique de données d'un établissement de recherche

Le projet ANR-BRIDGE a poursuivi la réflexion en analysant la pertinence de nouveaux principes directeurs (exigences, conformité, déontologie...) s'adressant à tout établissement de recherche encadrant ses activités de Gestion des Données de la Recherche.

C'est ainsi que BRIDGE propose deux nouveaux critères :

1. Un critère portant sur le volet « Légal et réglementaire ».
2. Un critère considérant le volet « Intégrité Scientifique » en recherche.

15. Légal et réglementaire

Le critère énonce les cadres légaux et réglementaires qu'il convient de respecter dans le cadre d'une politique de données d'établissement de recherche.

Panorama

Une politique de données doit préciser toute attente associée i) à la législation sur la **protection des données personnelles** en Europe et dans le monde ou ii) à une **législation contraignante**.

Citons pour exemples :

- Le règlement général sur la protection des données (**RGPD**)⁴⁸ est la loi la plus stricte au monde en matière de confidentialité et de sécurité. Bien qu'il ait été rédigé et adopté par l'Union

⁴⁸ Le règlement général sur la protection des données, <https://www.cnil.fr/fr/rgpd-de-quoi-parle-t-on>

européenne (UE), il impose des obligations aux organisations où qu'elles soient, tant qu'elles ciblent ou collectent des données liées à des personnes dans l'UE.

- Hors Europe⁴⁹, il est important de noter que différents niveaux de protection des données personnelles ont vu le jour, des pays se dotant d'une législation ou d'une autorité de protection des données, il convient de les consulter et de les respecter.
- Le Protocole de Nagoya⁵⁰ précise le cadre international du mécanisme d'*Accès aux ressources génétiques et de Partage juste et équitable des Avantages* découlant de leur utilisation (APA).

Dans le cadre de leurs activités de recherche, les scientifiques sont soumis aux **règles obligatoires issues des dispositions légales, conventionnelles et contractuelles** et indissociables des principes de la science ouverte.

Recommandations Bridge

R22. La politique indique clairement que le cadre légal et réglementaire des données fait partie de la gestion des données de la recherche. Elle fournit des liens vers des politiques connexes et/ou des documents d'appui. Des contacts sont identifiés.

16. Intégrité scientifique

Ce critère précise les dispositions de l'établissement en matière d'intégrité scientifique dans le cadre de la gestion des données de la recherche.

Panorama

"L'intégrité scientifique est la conduite intègre et honnête qui doit présider à toute recherche", Rapport Pierre Corvol, Professeur honoraire au Collège de France, 2016.⁵¹

Le Décret n° 2021-1572, déc. 2021⁵² relatif au respect des exigences de l'intégrité scientifique attend des établissements publics de recherche, une politique de conservation des résultats bruts de la recherche et de la mise à disposition des méthodes et protocoles, des données et des codes sources associés aux résultats de la recherche.

Les établissements de recherche :

- **Doivent former, sensibiliser et mettre en place le contrôle de l'intégrité scientifique.**
- **Sont incités à mettre en œuvre les plans de gestion des données.**

Recommandations Bridge

R23. La politique indique clairement que l'intégrité scientifique adresse la gestion des données de la recherche. Elle fournit des liens vers des politiques scientifiques connexes et des documents et dispositifs d'appui.

⁴⁹ <https://www.cnil.fr/fr/la-protection-des-donnees-dans-le-monde>

⁵⁰ Le protocole de Nagoya, <https://anr.fr/fr/lanr/engagements/le-protocole-de-nagoya/>

⁵¹ [Rapport Pierre Corvol](#), « Bilan et propositions de mise en œuvre de la charte nationale d'intégrité scientifique », 2016 remis au secrétaire d'État à l'enseignement supérieur et à la recherche.

⁵² <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000044411360>

VI. Conclusions et perspectives

En premier lieu, il nous est apparu qu'il n'existe pas de cadre harmonisé directement applicable pour l'élaboration d'une politique des données de la recherche, bien que des documents de cadrage aient été publiés au niveau national, européen et international.

La recherche scientifique évolue dans un environnement de plus en plus complexe et interdépendant, qui oblige les établissements, les collectifs scientifiques et d'appui à composer avec une multitude de variables d'alignement pour « rester conforme » avec :

- De nombreuses politiques et recommandations sur les données ;
- Des lois et des réglementations encadrant la donnée ;
- La propriété intellectuelle des résultats ;
- Une valorisation scientifique des travaux et des résultats de recherche ;
- Une pluralité d'acteurs et d'injonctions parfois contradictoires ;
- Les dispositifs et infrastructures numériques soutenant la recherche ;
- Une intégrité scientifique pour une recherche responsable.

Le contexte évolue et les travaux menés par le groupe d'intérêt RDA Data Policy contribuent à éclairer et à structurer ce paysage tant du côté des éditeurs de revues scientifiques que du côté des agences de financement.

Notre étude révèle que les travaux du groupe RDA Data Policy offrent l'opportunité pour les établissements de recherche de s'aligner sur un cadre de référence pour i) déchiffrer les complexités dans lesquelles la recherche scientifique s'inscrit, ii) parvenir à élaborer une politique de données qui s'interface avec les règles de cet écosystème à l'aune du mandat de chaque établissement de recherche et iii) rendre plus lisible une politique des données de la recherche auprès des chercheurs et des personnels de soutien.

Enfin, ce cadre de référence pourrait être mobilisé au sein des établissements de recherche, comme un instrument de sensibilisation, d'échange et de concertation auprès des services compétents dans les domaines adressés par la gestion des données de la recherche.

Gouvernance des données dans les établissements de recherche, rôles et responsabilités

Ce sujet central s'intéresse à la gouvernance de chaque établissement, dans sa mise en œuvre.

Dans la revue documentaire réalisée, les rôles et les responsabilités ne sont pas détaillés à ce niveau de gouvernance. Un travail pourrait consister, en s'appuyant sur le Plan national pour la science ouverte, à investiguer :

- les principes (scientifique, éthique, juridique, bonnes pratiques de gestion, innovation) appuyant la décision le long du cycle vie de la donnée et les acteurs (rôle et responsabilité individuelle, collective et institutionnelle) produits par INRAE dans un document de travail⁵³,
- les travaux initiés sur les rôles et responsabilités notamment par le groupe RDA SHARC IG, le projet FAIRsFAIR et les travaux sur le Data Stewardship menés au sein d'EOSC-Pillar⁵⁴.

Plans de gestion de données

Le plan de gestion de données (PGD) joue un rôle essentiel dans la planification de la gestion des données. Il nous paraît important de faire figurer le PGD dans les premiers critères d'un document de politique de données d'établissement de recherche.

⁵³ INRAE Dipso. Principes pour la gouvernance des données-Document de travail 2 INRAE version V9 - dec 2019

⁵⁴ <https://eosco-pillar.eu/data-stewardship-resources>

Celui-ci est exigé par les organismes de financement français et européens mais ce n'est pas une règle au niveau international ce qui génère des asymétries dans la conduite de projets hors Europe.

C'est une difficulté notée par les scientifiques et le personnel d'appui pour socialiser cette approche au sein d'un projet hors cadre européen, où le PGD n'est pas demandé. Le PGD relève plutôt d'une bonne pratique que chacun est libre de suivre, ce qui limite son utilité et son impact. Une sensibilisation sur l'intérêt de la mise en œuvre d'un PGD est à poursuivre. C'est une proposition de formation à l'international qui été formulée par le Cirad et l'IRD dans le cadre des ateliers de la donnée ECODOR⁵⁵ porté par l'Université de Montpellier.

Les partenaires du projet ANR-BRIDGE soulignent qu'il n'existe pas de règle établie sur la capitalisation ou le dépôt de la version finale d'un PGD dans leur établissement. Lors des ateliers du projet il a été précisé qu'à minima le PGD doit être partagé au sein du projet. S'il doit être diffusé, les partenaires suggèrent qu'il doit être déposé dans un entrepôt qui respecte les principes FAIR. Des archives ouvertes comme HAL ou Agritrop⁵⁶ au Cirad, acceptent désormais le type de document PGD. Le PGD devrait disposer d'un identifiant pérenne pour pouvoir s'y référer de façon non ambiguë. Cette approche permettrait d'appuyer à terme des besoins de suivi-évaluation, de capitalisation et de métriques des PGD.

S'approprier un cadre commun pour une politique des données des établissements de recherche

L'étude des politiques de données conduite par le groupe RDA Data Policy est axée sur la politique des éditeurs scientifiques. Nous avons analysé chacun de ces critères tels que présentés par le groupe RDA.

Ce cadre de référence pourrait être structuré avec la communauté de l'Enseignement Supérieur et de la Recherche en réaménageant l'ordre des seize critères pour composer un cadre approprié aux établissements de recherche. Enfin, il convient d'évaluer collectivement le cadre de référence de politique de gouvernance des données pour un établissement de recherche. Il sera en effet discuté lors de plusieurs conférences RDA France en 2022 et 2023.

Un des points d'attention est également de veiller à ne pas avoir des recommandations et des exigences discordantes au sein des politiques de données des tutelles qui travaillent en partenariat au sein d'Unités Mixtes de Recherche. La réflexion menée par les trois partenaires du projet ANR-BRIDGE permet dès à présent de formuler des propositions pour une meilleure coordination du partage des données de recherche notamment en UMR.

* * *

⁵⁵ <https://recherche.data.gouv.fr/fr/page/ateliers-de-la-donnee-des-services-generalistes-sur-tout-le-territoire>

⁵⁶ <https://agritrop.cirad.fr/cgi/search/advanced>