

Collège Données de la recherche - CoSO

« Pour une politique des données de la recherche :
guide stratégique à l'usage des établissements »

Date	Version	Statut	Contributeurs
18/01/2019	v. 0	Jalons	Séance plénière
20/02/2019	v. 1	Création du document	V. Stoll, PY Arnould
21/03/2019	v. 2.1	Ajouts, commentaires	C. Arènes, C. Sebban, T. Jouneau, J. Joanik, D. Chopard-Lallier
26/03/2019	v. 2.2	Ajouts, commentaires	Séance plénière
28/03/2019	v. 3	Synthèse	V. Stoll, PY Arnould
11/04/2019	v. 4.1	Ajouts, commentaires	C. Arènes, N. Couedel, C. Sebban, C. Espiau, T. Jouneau - synthèse V. Stoll, PY Arnould
16/04/2019	v. 4.2	Ajouts, commentaires	Séance plénière
18/04/2019	v.5	Synthèse	V. Stoll, PY Arnould
24/04/2019	v.5.1	Ajouts, commentaires	T. Jouneau, D. Chopard-Lallier, L. Albaret, S. Fortuno
24/04/2019	v5.2	Synthèse	V. Stoll, PY Arnould
12/11/2019	V5.3	Synthèse	V. Stoll, PY Arnould

Introduction

Le présent Guide identifie un socle de quelques objectifs fondamentaux à atteindre, complétés par des propositions d'actions additionnelles (recommandations), déclinables en fonction de l'ambition institutionnelle, des moyens et de la spécificité des unités de recherche.

Le Plan national pour la Science ouverte publié par le ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation le 4 juillet 2018 s'est donné comme ambition de rendre à terme les données issues de la recherche¹ financée sur fonds publics conformes aux principes FAIR² (axe n 2). Pour y parvenir, une série de mesures y est déclinée, notamment :

- rendre obligatoire la diffusion ouverte des données de la recherche issues de programmes financés par appels à projets sur fonds ;
- créer les conditions et promouvoir l'adoption d'une politique de données ouvertes associées aux articles publiés par les chercheurs ;
- construire autour de l'administrateur des données un réseau de correspondants dans les établissements, pour répondre aux questions que se posent les chercheurs sur les données de la recherche ;
- généraliser la mise en place de plans de gestion des données dans les appels à projets de recherche ;
- développer des centres de données thématiques et disciplinaires ;
- développer un service générique d'accueil et de diffusion des données simples.

Ce Guide s'efforce de tenir compte des différences existantes dans la gestion des données de la recherche, tant au niveau des disciplines que des établissements et opérateurs du MESRI.

Ce document ne traite pas des aspects liés aux publications³, à l'intégrité scientifique⁴, ni le volet juridique (licences, RGPD etc.), qui feront l'objet de travaux ultérieurs du CoSO.

¹ Selon l'OCDE, <http://www.oecd.org/fr/science/inno/38500823.pdf> : « 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 ».

² Acronyme pour *Findable, Accessible, Interoperable, Reusable*, soit *Facile à trouver, Accessible, Interopérable, Réutilisable*.

³ Se référer aux travaux du Collège publications.

⁴ Voir les travaux de l'Office français de l'intégrité scientifique : <https://www.hceres.fr/fr/ofis>.

Objectif 1 : Formaliser une politique d'établissement

Les enjeux de science ouverte deviennent de plus en plus prégnants dans les pratiques de recherche. Une des raisons invoquées réside en l'évolution des politiques des financeurs publics, qui glissent progressivement de l'incitation à l'obligation⁵. Plus largement, les objectifs majeurs d'une science ouverte concernent le partage des savoirs à des fins de validation et d'intégrité des pratiques scientifiques, l'encouragement à réaliser de nouvelles collaborations, la valorisation et la réutilisation des données produites lors de projets de recherche. Tous ces objectifs doivent en outre respecter le cadre de production établi lors des accords-cadres du projet et prendre en considération la réglementation en vigueur (loi 78-17 du 6 janvier 1978 modifiée, loi pour une République numérique du 7 octobre 2016 et règlement 2016/679 sur la protection des données).

Chaque opérateur de recherche à l'échelle nationale (université, établissement de recherche, regroupement...) est encouragé à se saisir de ces problématiques et à **adopter au plus haut niveau un texte-cadre**, définissant les grandes lignes politiques et son ambition en matière de science ouverte.

Recommandation n°1 :
Faire adopter un texte-cadre sur la science ouverte au niveau de l'établissement.

A minima, il est recommandé que ce document traite des sujets suivants⁶ :

- la position de l'établissement par rapport aux grands principes et déclarations fondateurs (Plan national pour la science ouverte, H2020/Horizon Europe...)⁷ ;
- la position de l'établissement sur le partage des données ;
- des recommandations sur les plans de gestion de données (PGD) pour les projets de recherche ;
- la sensibilisation et l'accompagnement des chercheurs et des équipes scientifiques grâce aux services opérationnels (services informatiques, services documentaires...) ;
- la formation des étudiants, doctorants, ingénieurs...

Selon l'engagement de l'établissement dans la gestion des données, une attention particulière peut être portée à :

- l'ouverture par défaut des données de la recherche qui peuvent l'être (*as open as possible, as closed as necessary*) ;

⁵ Voir notamment :

la politique européenne H2020 : <http://www.horizon2020.gouv.fr/cid82025/le-libre-acces-aux-publications-aux-donnees-recherche.html>

le Plan d'action 2019 de l'ANR : <http://www.agence-nationale-recherche.fr/financer-votre-projet/plan-d-action-2019/>

mais aussi à l'étranger la National Science Foundation : https://www.nsf.gov/news/special_reports/public_access/.

⁶ Pour mémoire, le dépôt des publications et l'intégrité scientifique ne doivent pas être omis.

⁷ Voir les exemples du CIRAD, de l'INRA ou de Paris-Nanterre.

- l'utilisation de modalités techniques facilitant le partage, telles que des identifiants pérennes, des formats et standards de structuration reconnus pour les jeux de données ;
- l'utilisation d'entrepôts certifiés institutionnels ou nationaux⁸.

1.1 : Faire adopter une politique d'établissement

L'adoption d'une politique d'établissement devrait être précédée d'une brève étude, qui pourra prendre la forme d'une enquête, analysant l'existant des pratiques et services, tout en identifiant les lacunes sur la gestion des données de la recherche et leur cycle de vie.

Il conviendra ensuite de bien déterminer le rôle des instances au niveau de la gouvernance, propres à décider ou inciter les actions à mener et à les diffuser, notamment :

- la présidence (pour une officialisation d'une politique en science ouverte et des plans de formation dans les écoles doctorales et masters) ;
- les conseils scientifiques, commission de la recherche et commission formation et vie universitaire (pour travailler avec les acteurs de la recherche et de la formation sur les contenus et la validation politique : président, référent intégrité scientifique) ;

Il est vivement recommandé de faire voter par les instances politiques (conseil d'administration, conseil scientifique) un texte-cadre, comme une charte de bonnes pratiques, reprenant par exemple⁹ :

- une déclaration d'intention ;
- les engagements de principe (signatures et adhésions aux principaux textes, comme l'Appel de Jussieu) ;
- la définition des objectifs cibles, remis dans le contexte des objectifs globaux (Plan national pour la science ouverte, politique de l'Union européenne...) ;
- les grandes lignes de la future feuille de route adoptée par l'établissement.

Selon les ambitions propres à chaque politique d'établissement, ce document-cadre pourra se limiter à une forte incitation ou au contraire se vouloir plus contraignant¹⁰.

Il est recommandé aux établissements de nommer un **chargé de mission science ouverte** dédié aux questions de science ouverte (archives ouvertes, données de la recherche...), afin d'accompagner la gouvernance de l'établissement.

Recommandation n°2 :
Nommer un chargé de mission science ouverte auprès de la gouvernance.

⁸ Voir notamment Cat OPIDoR : https://cat.opidor.fr/index.php/Cat_OPIDoR_wiki_des_services_d%C3%A9di%C3%A9s_aux_donn%C3%A9es_de_la_recherche.

⁹ Des exemples pourront être trouvés à l'INRA, au CIRAD, à l'Observatoire de Paris...

¹⁰ On pourra citer l'exemple de l'université de Marseille ou des Arts et Métiers Paris Tech.

1.2 : Encourager/valoriser la diffusion des données de la recherche des établissements

Les données de la recherche n'ont pas vocation à être produites et utilisées exclusivement pour un projet scientifique. Au contraire, une bonne gestion laisse entrevoir de nouvelles applications possibles, par exemple par le croisement ou la fouille de données. Pour ce faire, il est indispensable que les données soient stockées, structurées, interopérables et partagées (selon les principes FAIR), et, si possible, ouvertes (*open data*).

Plusieurs leviers peuvent être envisagés :

- encourager la **diffusion** et la **valorisation** des données et les faire reconnaître au niveau de l'établissement et des instances d'évaluations comme socle indispensable à la validation et l'intégrité scientifiques :
 - ✓ outre la prise en compte par les instances d'évaluation (HCERES, CN, CNU), les établissements peuvent se doter d'une politique incitatrice, par exemple par la prise en charge des frais de publications de jeux de données ou par la mise en œuvre d'un bonus de dotation pour des unités de recherche ou équipes exemplaires en matière de gestion de données ;
 - ✓ permettre la dissémination et la valorisation des données produites par l'établissement : entrepôt institutionnel de données de recherche référençant tout ou partie des données produites et les liant à l'archive institutionnelle pour les publications ; ou, si c'est possible, un référencement de tout ou partie des jeux de données produits dans l'archive institutionnelle¹¹ ;
 - ✓ encourager la publication de *Data papers* ;
- encourager le recours aux outils de fouilles de données, afin de permettre l'optimisation des champs de recherche et développer de nouvelles questions scientifiques ;
- encourager le recours aux outils de visualisation des données, intégrés ou non à un entrepôt institutionnel, propres à un sujet de recherche ou, idéalement, à plusieurs projets simultanément ;
- encourager la réutilisation des données, y compris au sein de l'établissement.

Recommandation n°3 :
Encourager l'écriture d'article de données (*Data papers*).

¹¹ Plus largement, impulser la création de centres de données thématiques et disciplinaires à des fins de recherche et de valorisation (exemple de plateformes ouvertes : <https://iscpif.fr/>)

Objectif 2 : Mettre en œuvre la politique d'établissement

Dès l'adoption des principes de la science ouverte par la gouvernance, les premières actions porteront d'abord sur la sensibilisation et le développement progressif de compétences et de bonnes pratiques.

2.1 : Faire adhérer l'ensemble de la communauté scientifique

L'adhésion progressive de l'ensemble des acteurs de la communauté scientifique (chercheurs, personnels d'accompagnement...) apparaît indispensable à une appropriation des enjeux.

Les bonnes pratiques existantes des équipes de recherches gagneront à être favorisées et valorisées. Leurs expériences concourront à alimenter les réflexions internes et à dynamiser les bonnes pratiques.

L'engagement des différents acteurs de la recherche (chercheurs, ITA, BIATSS...) dans la science ouverte et la gestion des données se doit d'être reconnu, y compris sur le plan de l'avancement des carrières.

Enfin, des opérations de sensibilisation et de formation permettront de diffuser une culture pratique de la science ouverte et de la science des données auprès de publics variés (masters, doctorants, mais aussi chercheurs et équipes d'appui à la recherche), en s'appuyant le cas échéant sur les acteurs locaux (MSH, URFIST...).

Recommandation n°4 :
Développer une culture de la science ouverte par la sensibilisation et la formation des différents acteurs de la recherche.

2.2 : Accompagner les pratiques de recherche

Il revient à chaque opérateur de recherche de définir ses propres processus internes et sa coordination d'ensemble pour la science ouverte (définition des rôles, dans un panel d'acteurs au profil diversifié).

Pour ce faire, il est nécessaire d'identifier les acteurs pivots pour un pilotage opérationnel optimal :

- la direction de la recherche et de la valorisation ;
- le collège des écoles doctorales ;
- les personnels de soutien à la recherche, notamment :
 - ✓ les bibliothèques, professionnels de l'IST, archivistes (formation des étudiants et équipes scientifiques et techniques, structuration/signalement/dépôt des données, connaissance des entrepôts) ;

- ✓ la direction du système d'information, du numérique (formatage des données, stockage, archivage) ;
- ✓ les services juridiques (former les étudiants et équipes scientifiques et techniques aux questions de propriété intellectuelle) ;
- ✓ le délégué à la protection des données (conformité RGPD) ;
- ✓ plus généralement, tout personnel ou toute unité de soutien ou service travaillant autour des données (infrastructures de recherche, MSH, centre de données...).

La coordination de ces différents acteurs pourra prendre par exemple la forme d'un guichet unique¹².

Recommandation n°5 :
Mise en œuvre d'un guichet unique.

Le rôle de ces acteurs gagnera en efficacité par le recrutement, ou à défaut la formation, de nouvelles compétences émergentes (*data manager, data curator, data scientist, data architect...*), qui devront faire l'objet d'une certaine reconnaissance de la communauté.

Pour coordonner et accompagner les bonnes pratiques d'une science ouverte dans les établissements, la création d'un poste d'administrateur de données de l'établissement devra être encouragée. Il aura pour fonction de mettre en œuvre la politique de la science ouverte des données de la recherche approuvée par la gouvernance. Il sera à même de coordonner l'action des différentes directions opérationnelles des établissements et du réseau d'appui à la recherche ou de son représentant. Il devra pouvoir disposer de compétences multidisciplinaires lui permettant de dialoguer avec les différents spécialistes cités ci-dessus, tout en s'appuyant sur les offres informatiques existantes¹³.

Recommandation n°6 :
Création d'un poste d'administrateur des données de l'établissement.

¹² Géré conjointement par plusieurs services (bibliothèques, directions de la recherche, direction informatique et services juridiques...), ce guichet est destiné à offrir un service global, homogène et transparent à l'utilisateur final. On peut imaginer, par exemple, un réseau d'appui à la recherche composé de correspondants "science ouverte" dédiés ou travaillant au sein de chaque unité de recherche, organisation permettant de mutualiser les expériences et développer une démarche d'amélioration continue. C'est une démarche adoptée avec succès par certaines universités à l'étranger : ainsi de celle d'Utrecht (<https://www.uu.nl/en/research/research-data-management>), d'Édimbourg (<https://www.ed.ac.uk/information-services/research-support/research-data-service>) ou Cambridge (<https://www.data.cam.ac.uk/>). En France, les Universités de Strasbourg ou de Lorraine s'orientent vers une organisation de ce type.

¹³ Par exemple, *data scientist, data curator, data librarian*, IT Dev, juristes... Se référer aux travaux du CoSO Collège compétences et formation.

2.3 : Mettre en place des outils de gestion de données (plan de gestion des données et démarche FAIR)

L'accompagnement des pratiques de recherche s'appuie sur des outils, comme les plans de gestion de données, qui reprennent toutes les étapes du cycle de vie des données de la recherche d'un projet¹⁴.

La **gestion des données de la recherche** consiste principalement à collecter et organiser les données, produits de la recherche, afin d'en faciliter l'accès et la réutilisation ultérieure et ce, dans un but final d'une utilisation efficiente des deniers publics. Les données de recherche de qualité forment donc un élément central du processus de recherche.

Une bonne gestion des données :

- décline autant que faire se peut les principes FAIR :
 - ✓ *Findable* :
 - a minima : préconiser l'utilisation d'identifiants pérennes, de type DOI,
 - et plus : préconiser l'utilisation de métadonnées standards ;
 - ✓ *Accessible* :
 - a minima : définir le régime par défaut d'ouverture et de réutilisation des données,
 - et plus : recommander une licence ; recommander l'usage d'entrepôts certifiés ;
 - ✓ *Interoperable* :
 - a minima : recommander l'adoption de formats interopérables ;
 - ✓ *Reusable* :
 - a minima : préciser la durée éventuelle d'un embargo,
 - et plus : recommander une licence ;
- s'appuie sur l'adoption d'un plan de gestion de données.

La rédaction d'un **plan de gestion des données**¹⁵ est devenue une obligation pour de nombreux bailleurs de fonds (Union européenne, ANR...). Ce document, qui peut être initialement très bref, évolue tout au long du projet. Il n'induit pas une ouverture des données par défaut (cas des données personnelles, sensibles, les brevets, etc.).

Ainsi, bien plus qu'une contrainte administrative, le plan de gestion de données se positionne comme partie intégrante du projet de recherche.

Quel que soit le modèle choisi, le plan de gestion de données doit traiter des aspects principaux du cycle de vie des données en suivant la ligne directrice des principes FAIR :

- la collecte des données (périmètre des données, méthodes utilisées, réutilisation de données existantes, etc.) ;
- le traitement des données, pendant et après le projet (description (métadonnées), structuration, validation des données et stockage provisoire) ;
- l'analyse des données (interprétation, production de résultats de la recherche, publications éventuelles) ;

¹⁴ Voir des exemples de cycle de vie des données :

<https://www.ukdataservice.ac.uk/manage-data/lifecycle>

<https://coop-ist.cirad.fr/actualites/integrer-la-gestion-des-donnees-aux-activites-de-recherche-poster>

¹⁵ Ou DMP, *Data Management Plan*.

- la conservation (migration vers des formats ouverts et supports appropriés, documentation, archivage) ;
- l'accès (partage, contrôle éventuel des accès, établissement de licences) ;
- la réutilisation ;
- des informations légales (propriété intellectuelle) et éthiques.

Recommandation n°7 :
Inciter l'écriture de plan de gestion de données dans tout projet de recherche.

Conclusion

Intégrer les données de la recherche à une réflexion plus globale sur la science ouverte, c'est avant tout mieux les gérer. Une politique ambitieuse des données de la recherche intègre les notions de reproductibilité et de réutilisation de la recherche, appuyées sur la qualité et la validation des jeux de données produits, qu'ils aient ou non vocation à être ouverts à la communauté scientifique.

C'est la prise en compte de l'ensemble des problématiques (juridiques, techniques, scientifiques...) touchant aux données de recherche qui permet, en favorisant leur ouverture, d'enrichir la recherche et de la rendre plus collaborative ; en préparant leur archivage, d'assurer la conservation et l'avenir de notre riche patrimoine scientifique ; enfin, en visant la transparence de la recherche et de ses résultats, de rendre la recherche intègre et responsable.

Liste indicative d'outils, services et référentiels :

Généralités

- Science Europe : <https://www.scienceeurope.org/>
- European research Council : <https://erc.europa.eu/>
- Research Data Alliance : <https://rd-alliance.org/>
- Digital Curation Center : <http://www.dcc.ac.uk/>
- Horizon Europe : <http://www.horizon2020.gouv.fr/>
- OPIDoR : <https://opidor.fr>
- DORANum : <https://dorum.fr/>
- Initiative GO FAIR data : <http://www.enseignementsup-recherche.gouv.fr/cid124728/science-ouverte-la-france-rejoint-go-fair-en-tant-que-co-fondatrice.html>
- European Open Science Cloud : <https://www.eosc-portal.eu/>
- Open Science Policy Platform : <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-policy-platform#>

Documents opérationnels

- Science Europe : https://www.scienceeurope.org/wp-content/uploads/2018/12/SE_RDM_Practical_Guide_Final.pdf
- European research Council : <https://erc.europa.eu/content/erc-data-management-plan-Template>
- Digital Curation Center : http://www.dcc.ac.uk/sites/default/files/documents/data-forum/documents/docs/DCC_Checklist_DMP_v3.pdf
- Horizon Europe : <http://www.horizon2020.gouv.fr/cid131357/en-route-vers-le-programme-horizon-europe-fp9.html>
- European Commission, *H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020*, 2016. Disponible sur http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- Guide d'analyse du cadre juridique en France sur l'ouverture des données de la recherche : https://inno3.fr/sites/default/files/2018-04/Guide_analyse_Cadre_Juridique_Ouverture_donnees_Recherche_V2_licenceOuvrte_prefaceDGRI.pdf
- Stratégie nationale des infrastructures de recherche (édition 2018, n°2) : <http://www.enseignementsup-recherche.gouv.fr/cid70554/la-feuille-de-route-nationale-des-infrastructures-de-recherche.html>