Recommandations aux revues souhaitant définir une “ politique de données ” liées aux publications

Romain Feret, Françoise Catherine Gouzi, Sandra Guigonis, Hélène Jouguet, Nicolas Larrousse, Armelle Thomas

▶ **To cite this version:**

Romain Feret, Françoise Catherine Gouzi, Sandra Guigonis, Hélène Jouguet, Nicolas Larrousse, et al.. Recommandations aux revues souhaitant définir une “ politique de données ” liées aux publications. [Rapport Technique] Comité pour la science ouverte. 2020, 8 p. hal-03594383

HAL Id: hal-03594383

https://hal-lara.archives-ouvertes.fr/hal-03594383

Submitted on 2 Mar 2022

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 aux revues souhaitant définir une « politique de données » liées aux publications

Mars 2021

Comité pour la science ouverte
Collège Données de la recherche

MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR,
DE LA RECHERCHE
ET DE L'INNOVATION
Liberté
Égalité
Fraternité

OUVRIR
LA SCIENCE !

CC BY

Recommandations aux revues souhaitant définir une « politique de données » liées aux publications

Ce document est mis à disposition des revues et des comités éditoriaux qui souhaitent mettre en place une *data policy* ou « politique de données ». Une politique de données définit ce que la revue attend de ses auteurs et autrices en matière de gestion et de diffusion des données liées à ses publications.

Ce document s'adresse en particulier aux éditeurs de revues en sciences humaines et sociales, dans la mesure où ils se sont relativement moins emparés de ce sujet que leurs homologues en sciences, technologies et médecine. Il peut cependant être utile à tous les éditeurs, quel que soit le périmètre disciplinaire de leur revue.

Les politiques de données sont de nature différente selon la nature des incitations et prescriptions qu'elles proposent, en particulier selon qu'elles :

- incitent ou obligent à diffuser tout ou partie des données qui sous-tendent les publications ;
- précisent des modalités de mise à disposition des données : délai, format, licence de partage… ;
- soumettent les données au processus de revue par les pairs au même titre que les publications.

Afin de mettre en place progressivement leur politique de données, les revues peuvent s'appuyer sur des typologies existantes (RDA propose par exemple 6 types de politique de données, Springer en définit 4).

Les données de la recherche comprennent l'ensemble « des documents se présentant sous forme numérique, autres que des publications scientifiques, qui sont recueillis ou produits au cours d'activités de recherche scientifique et utilisés comme éléments probants dans le processus de recherche, ou dont la communauté scientifique admet communément qu'ils sont nécessaires pour valider des conclusions et résultats de la recherche »[1].

Ce document est structuré en 7 sections et 4 colonnes :

- la 1re colonne comprend le nom de la section ;
- la 2e colonne est une description de la section présentée ;
- la 3e colonne précise les enjeux de la section et les questions que doivent se poser les revues ;
- la 4e colonne propose des exemples de formulation qui sont donnés **à titre indicatif**.

Ce document a été produit par le Collège Données de la recherche du Comité pour la science ouverte. Il est diffusé sous licence Creative Commons CC-BY et s'appuie notamment sur :

- Iain Hrynaszkiewicz, Natasha Simons, Azhar Hussain, Rebecca Grant, Simon Goudie. "Developing a Research Data Policy Framework for All Journals and Publishers". *Data Science Journal*, 19 (1). 2020. DOI: https://doi.org/10.5334/dsj-2020-005 ;
- l'adaptation qui en a été faite par l'Université de Toulouse-Jean Jaurès. Chloée Fabre, Françoise Gouzi. *Proposition de modèle de politique pour les revues et éditeurs quant aux données de la recherche*. 2020. ⟨hal-03026731⟩.

Citer ce document : Romain Féret[i], Françoise Gouzi[ii], Sandra Guigonis[iii], Hélène Jouguet[iv], Nicolas Larrousse[iv], Armelle Thomas[v]. *Recommandations aux revues souhaitant définir une « politique de données » liées aux publications*. Collège Donnée de la recherche du Comité pour la science ouverte. 2021.

[i] Normandie Université, Université de Lille

[ii] Université Toulouse – Jean Jaurès

[iii] Open Edition Center (UAR 2004), CNRS, Université d'Aix-Marseille, EHESS, Université d'Avignon

[iv] Huma-Num (UAR 3598), CNRS, Université d'Aix-Marseille, Campus Condorcet

[v] Maison des sciences de l'homme de Dijon (uB - CNRS)

[1] Directive (UE) 2019/1024 du Parlement européen et du Conseil du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public, s.l., 2019, vol.OJ L.

Section	Description	Enjeux et points à considérer	Exemples de formulation
1. Définition des données de recherche et exceptions	Décrit les données auxquelles la politique s'applique. Précise les exceptions éventuelles à cette politique.	**Enjeux** • Permettre aux auteurs de sélectionner les données concernées par la politique de la revue. • Permettre aux auteurs de préparer les données pour leur mise à disposition en lien avec la publication. **Points à considérer** • Quelles données sont concernées par la politique ? • Les données concernées devront-elles être les données brutes ou les données traitées qui ont sous-tendu les résultats présentés dans la publication ? • Les données contiennent-elles des contenus sensibles qui entrent dans le cadre du Règlement Général sur la Protection des Données (RGPD[2]) ? • Le cas échéant, quels traitements les données devraient-elles subir pour être conformes au RGPD ? • Quelles données font exception à la politique ?	*Cette politique s'applique aux données de la recherche qui seraient nécessaires pour vérifier les résultats présentés dans les publications de la revue.* *Les données de la recherche comprennent les données produites par les auteurs ainsi que les données provenant d'autres sources qui sont analysées par les auteurs dans leur étude. Ces données peuvent se présenter sous des formes variées : images, vidéos, tableaux de statistiques…* *Les données de la recherche qui ne sont pas nécessaires pour vérifier les résultats rapportés dans les contributions ne sont pas couvertes par cette politique.* *Cette politique sera 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.*
2. Standards et formats des données (et métadonnées)	Liste les standards principaux (et/ou les ressources où les trouver) utilisés pour les données et les métadonnées associées. Inclut nécessairement les protocoles de diffusion	**Enjeux** • Pouvoir trouver, lire et interpréter les données associées aux publications de manière durable. • Recommander l'utilisation de formats de fichiers ouverts et standardisés. • Sensibiliser les auteurs à l'importance et l'utilité du recours à des standards pour structurer des données et métadonnées.	*La revue incite les auteurs à privilégier les formats ouverts et standards. À titre d'exemple, la conformité des fichiers de données en vue d'un archivage pérenne auprès du CINES peut être vérifiée sur : https://facile.cines.fr* *Les métadonnées descriptives doivent être structurées en employant des standards reconnus, au minimum Dublin Core, disciplinaires ou plus*

[2] Le RGPD réglemente les modalités de traitement (collecte, traitement, conservation, diffusion…) des données à caractère personnel. Les délégués à la protection des données des établissements de recherche (ou DPO, *data protection officer*) accompagnent la mise en œuvre de ce règlement.

CC BY

	associés principalement aux métadonnées.	• S'appuyer sur les initiatives nationales et internationales existantes, telles que les groupes de travail de RDA : https://rd-alliance.org/groups/. **Points à considérer** • Existe-t-il des standards (*e.g.* structuration de métadonnées, vocabulaires, formats de fichiers, etc.) utilisés dans les disciplines couvertes par la revue ? Les métadonnées utilisées pour décrire le jeu de données lors de sa diffusion dépendent de l'entrepôt de données choisi (voir section 3). • Les jeux de données diffusés doivent au minimum être décrits par les métadonnées obligatoires du schéma Data Cite, à savoir : 1. Créateur du jeu de données, 2. Titre du jeu de données, 3. Éditeur ou hébergeur des données, 4. Année de publication, 5. Identifiant et son type (DOI, handle...), 6. Type de ressource.	*génériques. (https://fr.wikipedia.org/wiki/Dublin_Core).* *L'utilisation de vocabulaires dits « contrôlés », disciplinaires ou plus génériques, pour exprimer ces métadonnées est fortement recommandée (*e.g. *pour référencer un auteur https://orcid.org ; pour référencer un lieu https://www.geonames.org).*
3. Accès aux données et hébergement	Explique la manière dont les données devraient être hébergées pour que leur accès soit sécurisé et garanti sur le plus long terme possible. Précise si le dépôt des données dans un entrepôt spécifique est préconisé et, le cas échéant, ses	**Enjeux** • Assurer la préservation, la visibilité et l'accès aux données en les déposant dans un entrepôt. • Faciliter ainsi leur partage et leur réutilisation, et apporter les éléments constitutifs de la preuve scientifique. • Recommander l'utilisation d'un entrepôt qui va garantir la sécurisation des données et leur accessibilité sur le long terme. • Conseiller l'utilisation d'un entrepôt disciplinaire adapté à la revue. • Préciser les critères de choix d'un entrepôt, par exemple éviter l'utilisation d'un entrepôt de statut privé.	*Les données qui ont concouru à la rédaction de la contribution devront être déposées dans un entrepôt de données qui garantira la sécurisation du stockage et l'accès aux données, notamment par l'attribution d'un identifiant pérenne.* *Nous conseillons aux auteurs d'éviter l'utilisation des entrepôts de statut privé dont la feuille de route n'est pas transparente : modèle économique, gouvernance, pérennité… (ex. Figshare).*

	caractéristiques (*e.g.* certification, degré de conformité avec les principes FAIR, pertinence par rapport à la discipline concernée, etc.)	• Rappeler la mise en œuvre des principes FAIR dans la mise à disposition de ses données **Points à considérer** • La revue souhaite-t-elle suggérer un entrepôt en particulier ou laisser libre le choix de l'entrepôt ? • La revue couvre-t-elle un domaine disciplinaire pour lequel un entrepôt spécifique peut être désigné ? • Les auteurs doivent-ils utiliser un entrepôt certifié par CoreTrustSeal ? (voir les critères de certification et la liste des entrepôts certifiés sur le site web de CoreTrustSeal). • Si l'entrepôt n'est pas certifié, quels autres critères de sélection la revue préconise-t-elle ? - Reconnaissance par la communauté. - Attribution d'identifiants pérennes (DOI, handle…). - Licence de diffusion. - Lieu d'hébergement. - Préservation de long terme. - Statut public. - Formats de fichiers acceptés. - Formats des métadonnées. - Lien entre les données et la publication.	**Si la revue souhaite recommander un entrepôt particulier** *La revue recommande le dépôt des données dans l'entrepôt disciplinaire* [Nom de l'entrepôt] *(ex. Nakala en sciences humaines et sociales).* Dans ce cas, décrire l'entrepôt et le lien entre la revue et l'entrepôt : support proposé aux auteurs, collection de la revue sur l'entrepôt… **Si la revue souhaite faire des recommandations généralistes** *La revue recommande le dépôt des données dans un entrepôt qu'il soit généraliste (ex. Zenodo), institutionnel (ex. Data INRAE) ou disciplinaire (ex. beQuali pour les données d'enquête qualitatives).* *Dans tous les cas, les auteurs sont invités à vérifier que l'entrepôt choisi répond aux principaux critères de qualité suivants : https://doranum.fr/depot-entrepots/criteres-choix-entrepot/*
4. Modalités de mise à disposition des données	Explicite comment les données devront être mises à disposition et dans quel délai. Précise si les données font l'objet d'une évaluation par les pairs et sous quelles modalités.	**Enjeux** *Auteurs* • Avant la soumission, permettre aux auteurs de savoir à quel moment ils devront mettre leurs données à disposition du comité éditorial ou des relecteurs. • Informer que la licence choisie a des conséquences sur la capacité à réutiliser les données. Les revues ont intérêt à préconiser l'usage de licences ouvertes (ex. licences Creative Commons).	*Phase de soumission* *Les auteurs n'ont pas à transmettre leurs données lors de la soumission de leur contribution.* *Phase d'évaluation* *Si les relecteurs le jugent nécessaire, les auteurs doivent être en mesure de mettre à leur disposition*

		• Il existe des ressources pour savoir selon quelles modalités les auteurs peuvent diffuser leurs données : - Fiche DoraNum (Inist-CNRS) ; - Logigramme (INRAE). *Relecteurs* • Indiquer si une évaluation des données par les pairs est prévue ou requise et, quels en sont les critères. Le cas échéant, ils devront être accessibles de manière transparente sur le site de la revue. *Exemple de critères d'évaluation des données : https://zenodo.org/record/1065991#.X7T0MxJCcdU* • Préciser si les examinateurs sont également ou alternativement invités à évaluer si les données présentées sont conformes à la politique de la revue. **Points à considérer** • Est-ce que la revue souhaite accéder aux données dès la soumission de l'article, lors de sa révision ou uniquement pour préparer sa publication ? • Dans le cas où la diffusion des données avec l'article est obligatoire (sauf exceptions justifiées), un embargo est-il possible ? • Pendant combien de temps les auteurs sont-ils censés fournir un accès à leurs données ? • Sur quelle grille d'analyse, ou quels critères, est basée l'évaluation des données ? Cette grille est-elle transparente et accessible aux auteurs ?	*les données qui sous-tendent les résultats présentés dans leur contribution.* *Phase d'acceptation* *Les données devront être diffusées sans embargo ou après un délai le plus court possible, selon des modalités permettant leur réutilisation, avec un lien explicite entre les données et la publication qu'elles sous-tendent (voir sections 4 et 5).* *La revue encourage la mise à disposition des données sous des licences ouvertes qui permettent leur libre réutilisation. Les auteurs doivent utiliser les licences préconisées par l'entrepôt où les données ont été déposées.* *En publiant dans cette revue, les auteurs attestent de leur capacité à mettre les données et métadonnées à disposition du public pendant au moins 5 ans après la publication de leur contribution, par l'intermédiaire d'une plateforme ou par mise à disposition individuelle si les données ne peuvent être diffusées librement.* *Les alternatives à la diffusion en libre accès de données personnelles ou sensibles sont :* • *Anonymisation ou pseudonymisation des données avant diffusion en accès ouvert* • *Mise à disposition sur demande des données, uniquement à des fins de recherche* • *Diffusion uniquement des métadonnées relatives aux données*

5. Soutien aux auteurs et aux relecteurs	Décrit l'accompagnement que la revue propose aussi bien à destination de ses auteurs que de ses relecteurs.	**Enjeux** • Donner des informations sur les services auxquels les auteurs peuvent s'adresser pour obtenir une aide pour le partage des données de recherche. • Quand la revue prévoit une évaluation des données, fournir aux relecteurs des recommandations et procédures pour accéder, lire et comprendre les données qui accompagnent les publications. **Points à considérer** • Est-ce qu'un ou plusieurs des membres de la revue maîtrisent suffisamment le sujet pour répondre aux questions des auteurs ? des relecteurs ? • Quelles modalités d'accompagnement propose la revue : information en ligne, courriel dédié, ressources spécifiques ?	*Pour toute question relative à la politique de données de notre revue, les auteurs peuvent écrire à : revue.donnees[at]* *Nous invitons les auteurs à se rapprocher des services support de leur établissement en ce qui concerne les bonnes pratiques de gestion et diffusion des données.* *S'il existe un plan de gestion des données lié à ces données, les auteurs sont vivement invités à le consulter pour y trouver des réponses à leurs questions.*
6. Lien entre publications et données	Décrit les mécanismes d'association entre les données et les publications ainsi que les prérequis pour les réaliser comme par exemple l'affectation systématique d'identifiants pérennes.	**Enjeux** • Rendre visibles et accessibles les liens réciproques des données et des publications afin de pouvoir accéder aux données depuis la publication et inversement. • Faciliter la réutilisation des données. • Favoriser la confiance dans l'intégrité du processus scientifique et des méthodes engagées dans les travaux présentés. **Points à considérer** • Comment favoriser la citation des données et des publications et favoriser ainsi les liens croisés ? • Comment adopter une norme pour la citation des jeux de données ? Voir par exemple : https://datacite.org/cite-your-data.html	*Les auteurs sont invités à citer les jeux de données qui sous-tendent leurs publications dans une section spécifique « Données de recherche ». Cette section doit décrire les données disponibles, les modalités d'accès à ces données et fournir un lien pérenne pour y accéder.* *Cette section peut prendre une des formes suivantes ou une combinaison de celles-ci :* • *Les jeux de données générés et/ou analysés au cours de la présente étude sont disponibles dans l'entrepôt [Nom de l'entrepôt] ; [DOI].* • *Les jeux de données générés et/ou analysés au cours de la présente étude ne sont pas accessibles en accès ouvert en raison de*

			*[**spécifier la raison**] mais sont disponibles auprès de l'auteur sur demande justifiée.* • *Le partage de données ne s'applique pas à cet article car aucun jeu de données n'a été généré ou analysé au cours de la présente étude.* • *Les jeux de données sur lesquels s'appuie cette étude n'ont pas été créés par les auteurs. Ils sont disponibles en ligne : Créateur (Année de publication). Titre. Version. [Nom de l'entrepôt]. [DOI].*
7. Cas de non-respect de la politique	Décrit les risques encourus par les auteurs en cas de non-respect de la politique, afin d'éviter tout différend. Précise les procédures prévues en interne pour traiter ce genre de cas.	**Enjeux** • Prévoir les moyens d'assurer la mise en œuvre de la politique de la revue et éviter d'éventuels contentieux avec les auteurs. • S'assurer que les modalités de traitement de ces cas soient proportionnées et réalistes et que les décisions soient endossées par la revue en tant qu'entité et non par des personnes particulières. **Points à considérer** • La revue s'est-elle dotée de moyens pour repérer des cas de non-respect ou traite-t-elle uniquement les cas signalés par des lecteurs ou des relecteurs ? • Quels types de remédiations, voire de sanctions, sont envisagés : bandeau d'information sur l'article ? rétractation ? • Quelles procédures sont mises en place pour traiter ce type de cas ?	**Cas d'une politique incitative** *Cette politique a uniquement une valeur incitative. Aucune sanction n'est prévue dans le cas où les auteurs ne se conforment pas avec cette politique.* **Cas d'une politique coercitive** *Si une contribution ne respecte pas les règles prévues de mise à disposition des données, les auteurs en seront informés par le comité éditorial et ils disposeront d'un mois pour se mettre en conformité avec la politique de la revue.* *Si à l'issue de ce délai la contribution ne respecte toujours pas la politique de la revue, un bandeau sera ajouté en haut de l'article pour préciser en quoi la politique de la revue n'est pas respectée.* *Si l'incapacité à fournir les données remet en cause la crédibilité de la contribution ou qu'il est établi que certaines données sur lesquelles la contribution s'appuie sont erronées, la revue se réserve le droit de le rétracter.*