



HAL
open science

Politique des données des dispositifs et infrastructures de l'INEE

Cécile Callou, Isabelle Charpentier, Alyssa Clavreul, Hénon Amandine,
Dominique Joly, Emilie Lerigoleur, Mathieu Massaviol, Corinne Pardo

► To cite this version:

Cécile Callou, Isabelle Charpentier, Alyssa Clavreul, Hénon Amandine, Dominique Joly, et al.. Politique des données des dispositifs et infrastructures de l'INEE. 2022. hal-04000652

HAL Id: hal-04000652

<https://hal.science/hal-04000652v1>

Submitted on 22 Feb 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



Institut écologie et environnement

POLITIQUE DES DONNEES DES DISPOSITIFS ET INFRASTRUCTURES DE L'INEE

Le présent document constitue un cadre définissant les principes généraux d'utilisation, de stockage, de diffusion et de réutilisation des données. Il pourra évoluer pour permettre un alignement avec la politique institutionnelle du CNRS ou celle proposée plus largement aux niveaux national et européen.

Il s'adresse principalement aux producteurs et utilisateurs des données issues des dispositifs et infrastructures de l'Institut Écologie et Environnement (INEE) du CNRS. Les personnels des structures opérationnelles de recherche multi-tutelles peuvent être amenés à appliquer d'autres recommandations, en fonction de leur implantation.

Les principes énoncés suivent ceux de la politique de Science Ouverte : "aussi ouvert que possible, fermé autant que nécessaire". Lorsque les conditions d'ouverture des données ne peuvent être remplies pour des motifs d'ordre juridique ou de confidentialité (par exemple en cas de présence de données sensibles ou personnelles), ces conditions doivent être clairement établies.

1. Périmètre et application

Cette politique s'applique aux données produites dans le cadre des :

- [Zones Ateliers](#) (ZA) et leur réseau (RZA) qui se focalisent autour d'unités fonctionnelles (un fleuve et son bassin versant, les paysages - agricole ou urbain - et la biodiversité, de l'antarctique à l'Afrique subsaharienne, ou le littoral, ou encore les environnements caractérisés par une irradiation chronique d'origine naturelle ou naturelle renforcée) et y développent une démarche scientifique spécifique en s'appuyant sur des observations et expérimentations sur des sites ateliers, pour y mener des recherches inter- et transdisciplinaires sur le long terme ;
- [Observatoires Hommes-Milieu \(OHM\)](#) et du LabEx qui les fédère ([DRIHM](#) - Dispositif de Recherche Interdisciplinaire sur les Interactions Hommes-Milieu) ; ils se consacrent, par la convergence de toutes les Sciences de l'Environnement (SDE), à l'étude des socio-écosystèmes anthropisés affectés par des crises d'origine anthropique ;
- [AnaEE France](#) qui offre à la communauté scientifique un accès complet à l'étude des écosystèmes terrestres et aquatiques, au travers de ses plateformes expérimentales distribuées sur l'ensemble du territoire national (métropole, départements et régions d'outre-mer - DROM-COM) et de ses Écotrons (Montpellier et Centre de Recherche en Écologie Expérimentale et Prédicative de Saint Pierre-lès-Nemours),

- [Stations d'écologie expérimentales](#) et leur réseau RenSEE qui sont ancrés dans une dynamique de temps long ; elles sont installées au cœur de milieux naturels qui couvrent la plupart des biotopes, tempérés et tropicaux, montagnards, ainsi que terrestres et marins qui ont pour vocation de développer une large gamme d'instruments et de protocoles expérimentaux.
- [Sites d'Étude en Écologie Globale](#) (SEEG) qui sont des lieux d'étude de taille variable où la recherche se concentre sur un unique sujet bien identifié et se développe en écologie globale, c'est-à-dire en interdisciplinarité des SDE, tout le spectre n'étant pas cependant nécessairement impliqué.

Elle concerne l'ensemble des personnels scientifiques (chercheur·e·s, enseignant·e·s-chercheur·e·s, doctorant·e·s, ingénieur·e·s et technicien·ne·s) impliqués dans les travaux de recherche menés au sein de ces plateformes expérimentales ou de ces observatoires. Dans le cas de projets de recherche faisant l'objet de conventions spécifiques, les porteurs de projet s'attacheront à respecter les principes de cette politique en accord avec les organismes partenaires.

Pour rappel, les données de recherche sont définies comme des "enregistrements factuels (chiffres, textes, images, sons, etc.), 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 recherche¹".

Les publications scientifiques ne se substituent pas au jeu de données publié. Elles ne traitent en général que d'un sous-ensemble du jeu de données et ne fournissent pas toujours l'intégralité des informations nécessaires à la réutilisation des données. Il conviendra en particulier d'indiquer si les données ont été traitées ou transformées².

2. Politique d'ouverture des données : la loi et les recommandations aux infrastructures de recherche dans le contexte de la Science Ouverte

En France, l'ouverture par principe des données publiques disponibles au format électronique est posée par la loi pour une République numérique 2016-1321 du 7 octobre 2016 ([Loi « Lemaire »](#)). Cela concerne les données des organismes publics de recherche et celles dont la publication présente un intérêt économique, social, sanitaire ou environnemental. De même, la loi prévoit un principe de libre réutilisation de ces données.

¹ Définition de l'Organisation de Coopération et de Développement Économique (OCDE)

² Dans leur cycle de vie, les données peuvent faire l'objet de transformations. On distingue ainsi les :

- données brutes ou primaires (données d'observation, tensions délivrées par un capteur, par exemple). Elles expriment un état objectif du réel, en dehors de toute interprétation arbitraire ou transformation ;
- données transformées. Il s'agit de données primaires compilées, agrégées, synthétisées. Elles intègrent l'ensemble des procédures de filtrage et d'étalonnage mis en œuvre sur les données délivrées (filtrage automatique et manuel, étalonnage externe, correction des facteurs externes influençant la mesure) ;
- données élaborées par traitement pour arriver à un résultat consolidé, données interprétées/analysées en réponse à une problématique précise, ainsi que les informations complexes (données corrélées, indices, indicateurs, etc.). Les traitements peuvent être :
 - des techniques permettant de combler les données manquantes ;
 - des agrégations de capteurs ;
 - l'assimilation de données dans un modèle ;
 - la reconstitution d'une donnée non directement mesurée.



En particulier, « dès lors que les données issues d'une activité de recherche financée au moins pour moitié par des dotations de l'État, des collectivités territoriales, des établissements publics, des subventions d'agences de financement nationales ou par des fonds de l'Union européenne ne sont pas protégées par un droit spécifique ou une réglementation particulière [...] leur réutilisation est libre (Art. L. 533-4) ». Enfin, toute diffusion des données doit « se faire dans un standard ouvert, aisément réutilisable et exploitable par un système de traitement automatisé ». L'ouverture par principe des données publiques de recherche et donc leur libre réutilisation ne pourra être limitée qu'en cas d'exceptions strictement encadrées par la loi³.

La France s'est depuis dotée de Plans Nationaux pour la Science Ouverte ([PNSO juillet 2018](#) et [PNSO2 juillet 2021](#)), qui réaffirment les principes d'ouverture des résultats de la recherche, et d'un décret relatif au respect des exigences de l'intégrité scientifique par les établissements publics contribuant au service public de la recherche et les fondations reconnues d'utilité publique ayant pour activité principale la recherche publique ([Décret n° 2021-1572 du 3 décembre 2021](#)). À ce titre, les établissements publics « promeuvent la diffusion des publications en accès ouvert et la mise à disposition des méthodes et protocoles, des données et des codes sources associés aux résultats de la recherche afin d'en garantir la traçabilité et la reproductibilité (art.2) ».

Des délais d'ouverture des données peuvent cependant se justifier au regard du temps de publication des résultats scientifiques, des pratiques de la communauté scientifique concernée, des besoins des futurs utilisateurs des jeux de données ou encore de questions éthiques.

C'est dans ce contexte législatif que le CNRS a présenté une [feuille de route pour la Science Ouverte](#) (novembre 2019), complétée par un [plan Données de la recherche](#) (novembre 2020) que les structures opérationnelles de recherche, les dispositifs et les infrastructures doivent mettre en œuvre.

3. Principes de stockage, de diffusion et de (ré)utilisation des données

L'Institut Écologie et Environnement (INEE) du CNRS recommande aux producteurs des jeux de données de :

- **publier leurs données dans un entrepôt de données** préférentiellement institutionnel ou disciplinaire sécurisé, voire certifié, tel que [data.indores](#) (CNRS-INEE), et [Nakala](#) (TGIR Huma-Num) pour les données spécifiques Sciences Humaines et Sociales. Des identifiants pérennes⁴ seront assignés aux données.
- **renseigner les métadonnées**⁵. Celles permettant de rendre les données “facilement trouvables” devront être fournies dans tous les cas.

³ Les exceptions sont :

- le secret en matière commerciale et industrielle et le secret professionnel ;
- le secret de la défense nationale et des impératifs relatifs à la sécurité de l'État ou de l'établissement (ex. sécurité des systèmes d'information des administrations), PPST ;
- les droits de tiers ;
- les documents non encore librement communicables au regard du code du patrimoine.

⁴ Les identifiants pérennes garantissent un lien stable à la ressource en ligne. Ils font correspondre en permanence l'identité de la ressource à sa localisation sur le web. Il existe deux grands types : les identifiants pérennes “objet” (pour les publications et données) et “contributeur” (pour les auteurs et institutions). Source : [Dorandum](#)

⁵ Les métadonnées sont des données descriptives du jeu de données : auteur(s), date de création, titre/résumé du jeu de données, format, emprise spatio-temporelle, contacts, droits d'accès/utilisation,



- **structurer les données** en conformité avec les principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable⁶).
- **anticiper les conditions juridiques** d'utilisation des données après la fin du projet de recherche en choisissant des licences explicitant les droits attachés aux données ([Etalab](#) ou [CC BY](#) par exemple).
- **discuter de toute question légale ou éthique particulière** avec le Délégué à la Protection des Données (DPD) de l'établissement employeur.

Il est à noter que la question de l'archivage des données n'est pas traitée dans ce document⁷.

L'Institut Écologie et Environnement du CNRS (INEE) demande aux utilisateurs des jeux de données publiés dans ces dispositifs et infrastructures de :

- **vérifier les conditions d'accès et d'utilisation de la donnée, notamment en prenant connaissance de la licence associée.**
- **comme pour les publications scientifiques, respecter les principes de citation** en se référant aux informations disponibles sur l'entrepôt.



Roscian, Marjorie; Herrel, Anthony; Zaharias, Paul; Cornette, Raphaël; Fernandez, Vincent; Kruta, Isabelle; Cherel, Yves; Rouget, Isabelle, 2022, "Replication Data for: "Every hooked beak is maintained by prey: ecological signal in cephalopod beak shape"", <https://doi.org/10.48579/PRO/BVWXJL>, data.InDoRES, V1

Citer le jeu de données ▾

Pour en savoir plus, consultez le document [Data Citation Standards](#).

Les dispositifs et infrastructures portés ou co-portés par l'Institut Écologie et Environnement du CNRS suivent les principes mentionnés précédemment. En fonction de leur périmètre d'intervention, ils peuvent être amenés à préciser certaines modalités. Voir par exemple la [politique des données d'AnaEE-France](#).

Le présent document a été validé par la DDOR en juillet 2022.

Pour toute information se rapportant à ce document : [représentant INEE de la DDOR](#)

Pour toute information sur les données de la recherche à l'INEE : contact@indores.fr

dictionnaire des variables, protocoles associés (qui, quoi, comment, où, etc.).... Elles sont nécessaires à la recherche et à la compréhension des données et à leur exploitation.

⁶ Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016)

⁷ Sur la distinction entre stockage et archivage pérenne : <https://dorum.fr/stockage-archivage>

Historique des versions

Date	Version	Auteurs	Descriptif
04/07/2022	V1.0	Callou C., Charpentier I., Clavreul A., Hénon A., Joly D., Lerigoleur E. Massaviol M., Pardo C. Relecture par Sylvie Rousset (DDOR)	Création du document et relecture et validation par la DDOR du CNRS

