



INEE

Politique des données des entités de recherche

Webinaire RDA-France

l'exemple de l'Institut Ecologie et Environnement

26 mai 2023

Cécile Callou

UAR BBEES – Bases de données sur la Biodiversité,
Ecologie, Environnement et Sociétés

CNRS-MNHN

Quelle est l'entité pour laquelle la politique des données a été élaborée?

A l'origine

Proposer aux unités INEE une déclinaison **adaptée** de la feuille de route du CNRS pour la Science ouverte (2019) et du Plan données de la recherche du CNRS (2020)



Qui a piloté l'élaboration de la politique?

Début 2022 - réflexion et rédaction par un petit groupe de travail composé de :

- ❖ Dominique Joly , Dir. Adj. Scient. **Grands équipements et Infrastructures (Ecotrons, AnaEE, ReNSEE, Zones Ateliers)**
- ❖ **BBEES** : Cécile Callou (dir. unité) & Amandine Henon (projet InDoRES)
- ❖ **AnaEE France** : Alyssa Clavreul (chef projet ST)
- ❖ **Réseau Zones Ateliers** : Isabelle Charpentier (resp.)
- ❖ **DRIIHM** (Dispositif de recherche interdisciplinaire sur les interactions hommes-milieus) : Corinne Pardo (resp.) & Mathieu Massaviol (chargé projet)
- ❖ **RZA & DRIIHM** : Emilie Lerigoleur

Point de départ de la réflexion : travail effectué dans le cadre d'AnaEE-France

Point de départ de la rédaction : valider une politique générale de type « tronc commun INEE » avec une déclinaison par grands dispositifs

Qu'est-ce qui a marché?

Production d'un texte de 4 pages, en ligne sur site de l'INEE et sur HAL Relu et validé par la DDOR du CNRS le 4 juillet 2022



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 Ecologie et Environnement (INEE) du CNRS. Les personnels des structures opérationnelles de recherche multi-tâches 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 Milieux (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'Arctique subarctique, ou le littoral, ou encore les environnements caractérisés par une transition chimique d'origine naturelle ou naturelle renforcée) et y développent une démarche scientifique spécifique en s'appuyant sur des observations et expérimentations sur des sites atelers, pour y mener des recherches inter- et transdisciplinaires sur le long terme ;
- **Observatoires Hommes Milieux (OHM)** et du LabEx qui le fédère (**OHM-L**) ; Dispositif de Recherche Interdisciplinaire sur les Interactions Hommes Milieux ; Il se concentre, 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 ;
- **Amélior 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 Écosétoles (Montpellier) et Centre de Recherche en Écologie Expérimentale et Prédictive de Saint Pierre-Mémoris).

- **Plateaux d'écologie expérimentale**, et leur réseau RenSÉE 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 de l'étude en écologie globale (EEG)** 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'attachent 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 recherche".

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 conviendrait en particulier d'indiquer si les données ont été traitées ou transformées¹.

2. Politique d'ouverture des données : la loi et les recommandations aux infrastructures de recherche dans le cadre 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 ([voir le lien](#)). 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 :
• données brutes ou primaires (données d'observation, tensions délivrées par un capteur, par exemple). Elles équipent un état objectif du réel, en dehors de toute interprétation arbitraire ou transformation ;
• données transformées... Il s'agit de données primaires complètes, agrégées, synthétisées. Elles intègrent l'ensemble des procédures de filtrage et d'échantonnage mis en œuvre sur les données délivrées (filtrage automatique et manuel, échantonnage 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 complémentaires (données connexes, indices, indicateurs, etc.). Les traitements peuvent être :
• des techniques permettant de combiner les données manquantes ;
• des agrégations de capteurs ;
• des rassemblements de données dans un modèle ;
• la reconstruction 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 loi².

La France s'est depuis dotée de Plans Nationaux pour la Science Ouverte (**PNSO juillet 2011** et **PNSO 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° 2011-1317 du 3 décembre 2011**). 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 détails 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) qui des 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 Ecologie 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, libre certifié, tel que **data4science** (CNRS-INEE), et **Hal** (IGDR-Huma-Norm) pour les données spécifiques Sciences Humaines et Sociales. Des identifiants pérennes³ seront assignés aux données.
- **renseigner les métadonnées** (elles 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), PPGT ;
• 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 "DOI" (pour les publications et données) et "contributor" (pour les auteurs et institutions). Source : [Openaire](#).
⁵ Les métadonnées sont des données descriptives du jeu de données : auteurs, 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, Interprétable, Réutilisable)⁴ ;
- **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 ([Creative Commons](#) ou [CC-BY](#) par exemple) ;
- **discuter de toute question légale ou éthique particulière** avec le Délégué à la Protection des Données (DPO) de l'établissement employeur.

Il est à noter que la question de l'archivage des données n'est pas traitée dans ce document⁵.

L'Institut Ecologie 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.



Les dispositifs et infrastructures portés ou co-portés par l'Institut Ecologie 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 le [politique des données d'Amélior France](#).

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

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

Pour toute information sur les données de la recherche à l'INEE : [contact@ecologie.fr](#)

diccionaire des variables, protocoles associées (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://forum.fairsharing.org/stockage-archivage](#)

Quelles difficultés avez-vous rencontré? Qu'est-ce qui pourrait faciliter ce processus?

Question non triviale : jusqu'où aller ?

- ❖ Choix a été fait de ne s'intéresser qu'aux dispositifs et infrastructures de l'INEE, évitant ainsi la question des cas de multi-tutelles.
- ❖ Rappel des lois et recommandations dans le contexte de la Science ouverte
- ❖ **Préconisations** pour le stockage, la diffusion et la (ré)utilisation des données

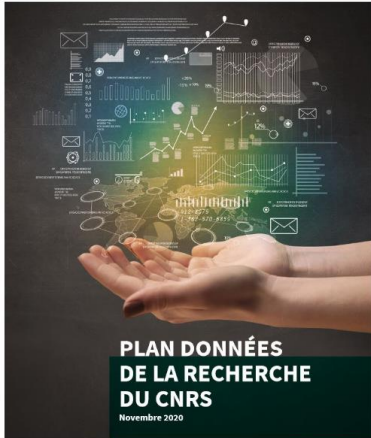
Diffusion du texte

- ❖ À qui ?
- ❖ Comment ? (rôle de la DDOR)

→ attente d'un élargissement s'appliquant à tous les agents de l'INEE

Documents de référence utilisés ou produits

<https://www.anaee-france.fr/ressources/donnees-anaee-france/324-politique-des-donnees-anaee-france>



Plan Données de la recherche du CNRS
[Télécharger \(novembre 2020\)](#)



Feuille de route science ouverte du CNRS
[Télécharger \(novembre 2019\)](#)

<https://www.science-ouverte.cnrs.fr/>

Politique des données AnaEE-France

Ce document définit la politique des données acquises à partir des services de l'infrastructure nationale AnaEE-France et leur modalité de diffusion. Nous rappelons que ces services couvrent l'accès aux plateformes expérimentales, la production d'analyses pour caractériser les écosystèmes, la mise à disposition d'instruments et l'accès aux plateformes de modélisation. Ce document ne traite pas des données générées par des modèles utilisant les plateformes de modélisation qui font l'objet d'un document spécifique.

Ce document constitue un cadre définissant les **principes d'utilisation et de diffusion des données** qui devra trouver par la suite une traduction juridique dans les documents appropriés (licences, chartes) et s'appuiera sur des outils informatiques de mise à disposition des données. Ce document met à jour les principes énoncés dans l'accord de consortium AnaEE-France. Ce document devra évoluer pour permettre un alignement avec la politique qui sera décidée au niveau Européen tout en restant conforme à la législation Française et aux politiques institutionnelles (CNRS, INRA et autres tutelles des plateformes AnaEE-France).

Métadonnées caractérisant les projets d'expérimentation

Toutes les demandes d'utilisation des services font l'objet d'une rédaction de projet définissant la conception de l'expérience et les ressources à mobiliser. Le projet soumis via une interface de soumission de projet permet de collecter des métadonnées qui seront visibles, sans restriction, sur le site AnaEE-FR, dès la fin du projet. Les métadonnées « projet » comportent :

- Titre et acronyme*
- Les services AnaEE-France mobilisés*
- Le projet de recherche (utile pour associer des projets d'expérimentation mobilisant plusieurs services)
- Un descriptif résumé du projet : objectifs, conception de l'expérience ou des expériences réalisées, résultats attendus
- Un résumé grand public (optionnel) *
- Le type d'écosystème
- Les variables mesurées
- Les facteurs manipulés
- Dates d'exécution sur la plateforme et d'exécution du projet
- Responsable du projet* (attention à bien respecter les clauses RGPD en ajoutant une note de clauses standard et un lien à cliquer par l'utilisateur)
- Liens vers les publications (à actualiser au fil de l'eau)

* champs diffusés dès l'acceptation du projet sous réserve d'accord du responsable du projet.