



# PLAN DE GESTION DES DONNÉES DU RÉSEAU DES ZONES ATELIERS

Novembre 2020

Auteurs et Affiliations

V. Girard (CNRS)

<b>Version</b>	<b>Statut / Modifications apportés</b>	<b>Date</b>	<b>Auteurs</b>
Version 1.0	Draft	23 nov. 2020	V. Girard
	Relecture	12 janv. 2021	D. Joly
	Relecture	27 jan. 2021	A. Henon

# SOMMAIRE

PREAMBULE	3
0. Aperçu général	4
<b>1. Description des données et collecte ou réutilisation de données existantes</b>	<b>4</b>
2. Documentation et qualité des données	5
3. Stockage et sauvegarde pendant le processus de recherche	7
4. Exigences légales et éthiques, codes de conduite	7
5. Partage des données et conservation à long terme	8
6. Responsabilités et ressources en matière de gestion des données	8
<b>ANNEXES</b>	<b>10</b>

# PREAMBULE

Le Réseau des Zones Ateliers (RZA) est devenue Infrastructure de Recherche<sup>1</sup> des socio-écosystèmes en mai 2018 et compte 14 Zones Ateliers (ZAs), labellisées pour les plus anciennes, il y a près de 20 ans. Le RZA est conçu comme une opération de recherche qui consiste à décrire, comprendre et prédire la réponse d'écosystèmes plus ou moins anthropisés au changement global, pour formaliser et théoriser le fonctionnement des socio-écosystèmes, et aider ainsi à leur gestion et leur gouvernance. La démarche intègre une communauté d'acteurs incluant les chercheurs au sein d'un réseau beaucoup plus large associant tous les acteurs de la société civile. Cette communauté innove et expérimente pour générer des socio-écosystèmes durables, pilotés selon des principes de gestion adaptative et par une gouvernance elle-même adaptative. Un schéma conceptuel partagé a permis d'harmoniser les questions, les protocoles et les données collectées (à la fois pour des variables écologiques et socio-économiques ; Bretagnolle et al. 2019<sup>2</sup>).

La gestion de la donnée au sein du RZA s'organise autour de la centralisation des métadonnées à des fins d'interfaçage avec les portails nationaux (PNDB, Geoportail<sup>3</sup>, etc.) et européens (Elter) et des actions transversales thématiques. Notamment, le RZA s'appuie sur le collectif BED<sup>4</sup> pour développer des outils (ex. Geoflow) et prendre à charge l'animation de gestion de la donnée et la formation des membres des ZAs (ex. ET E-ENVIR 2021).

Les informations qui suivent visent à offrir une vision synthétique du fonctionnement, des acteurs et moyens mis en jeu, avec une diffusion auprès de nos partenaires, notamment institutionnels.

---

<sup>1</sup> [https://data.enseignementsup-recherche.gouv.fr/pages/infrastructure-de-recherche-fdr2018/?refine.acronyme=E - LTER-France%20RZA](https://data.enseignementsup-recherche.gouv.fr/pages/infrastructure-de-recherche-fdr2018/?refine.acronyme=E-LTER-France%20RZA)

<sup>2</sup> Bretagnolle V., Benoit M., Bonnefond M., Breton V., Church J., Gaba S., ... & Hervé Fritz (2019) Action-orientated research and framework: insights from the French Long-Term Social-Ecological Research network. *Ecology & Society*, 2019.

<sup>3</sup> <https://www.geoportail.gouv.fr/>

<sup>4</sup> <http://www.za-inee.org/fr/data-md-intro>

## 0. Aperçu général

Le présent document vise à décrire les éléments fonctionnels et organisationnels de la gestion des données du RZA.

La gestion de la donnée au sein du RZA s'organise autour de la centralisation des métadonnées à des fins d'interfaçage avec les portails nationaux (PNDB, Géoportail<sup>5</sup>, etc.) et européens (eLTER) et la production et le partage de jeu de données au sein des actions transversales thématiques.

L'IR RZA rassemble 14 Zones Atelier (ZAs), plateforme d'observations des socio-écosystèmes sur le long terme, distribuées en France Métropolitaine (12 ZAs), en Antarctique et Afrique du Sud/Zimbabwe. Les ZAs documentent les compartiments du capital humain et social, et du capital biophysique, et les interactions associées entre les deux (Bretagnolle et al. 2019). Au sein de chaque ZA on peut distinguer des sites ateliers plus ou moins instrumentés dont la taille varie de quelques ha à plusieurs km<sup>2</sup>. Chaque ZA et site atelier diffèrent dans leurs objectifs initiaux, expliquant l'hétérogénéité des choix de capteurs, des protocoles mis en œuvre et de la gestion des données.

Chaque ZA dispose de son propre modèle de gestion des données et fait appel soit à des ressources en interne soit à des ressources externes (géocatalogue, BDD, serveurs, ressources humaines, etc.; à l'échelle locale et/ou nationale). Les principaux appuis se trouvent auprès des OSU, UMS BBEES et parfois des partenaires non académiques.

## 1. Description des données et collecte ou réutilisation de données existantes

### 1a. ACQUISITION DES NOUVELLES DONNÉES

*1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?*

Les données collectées au sein des ZAs sont acquises par l'intermédiaire de capteurs in-situ, de satellites, de relevés, d'analyses en laboratoires, et d'enquêtes, à des pas de temps allant de la minute à une fois par an, voire à des fréquences pluriannuelles et peuvent être produites à partir des jeux et bases de données produites par des ZAs (compilation des données). **A PRECISER**

Les protocoles mis en œuvre au sein des ZAs sont partagés par l'intermédiaire des fiches de métadonnées présentant les jeux ou bases de données d'intérêt (item "provenance") et parfois directement accessible en ligne ou à travers les publications (cf. Plan de gestion des données des ZAs; **Annexe X**).

L'utilisation de données préexistantes est conditionnée aux termes des accords entre les ZAs et leurs partenaires, entre les laboratoires au sein des ZAs. En ce qui concerne le partage des données dans les actions transversales, seules les données disposant d'une levée d'embargo et d'un libre droit d'accès peuvent être prises en compte. La provenance des données sera documentée en associant les DOI des fiches de métadonnées et/ou des jeux de données déposés.

Le principal outil de gestion à l'échelle du RZA est geoflow (REF).

### 1b. TYPE ET FORMAT DES DONNÉES

*1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?*

Les données collectées sont actuellement des données numériques (bases de données, tableurs), textuelles, des rasters, des images, des vidéos et du son; qui répondront aux exigences d'utilisation de l'entrepôt cible (cf. tableau ci-après). Ces données ont vocation à évoluer au regard d'acquisitions de capteurs et de méthodes de mesure continue.

**Tableau avec types de données et format associés**

Le volume des données est compris entre 10 et 100 Go de données. Toutefois, aucune information précise sur le volume des données est connue (chantier de recensement en cours). Cette volumétrie sera amenée à augmenter au fur et à mesure des années et devra être revue régulièrement.

---

<sup>5</sup> <https://www.geoportail.gouv.fr/>

Un format standardisé des données sera donné aux producteurs de données pour faciliter les opérations de partage; notamment il s'agira d'avoir des indexations de fichiers de données et un ensemble de fichiers associés aux variables partagées. A COMPLETER en discussion avec BBEES (Darwin Core ou proche, pour alléger la mise en forme).

*A compléter au fil des dépôts sur le Data-InDoRES dataverse de BBEES. Reprendre la liste des publis et/ou fiche de métadonnées indiquant des DOI.*

## 2. Documentation et qualité des données

### 2a. CENTRE DE RESSOURCES

L'ensemble des informations à destination de la communauté du RZA est accessible sur la plateforme collaborative [POPS](#), hébergée et mise à jour par l'UMR 6554 LETG du CNRS en partenariat avec l'OSU IUEM et la société Dotgee ([détails sur POPS](#)).

Le lien associé au projet BED est : <https://www-iuem.univ-brest.fr/pops/projects/za-bancarisation-bed>

Les personnes en charge de l'administration du projet sur POPS sont : Christine Plumejeaud, Cécile Pignol, Virginie Girard et Isabelle Charpentier. La liste fédère 89 personnes, dont 44 contributeurs et 41 observateurs.

L'ensemble des documents partagés est destiné à un large public et sont des documents amendés. La plateforme collaborative offre un suivi des activités, des annonces et un espace Ressources qui se divise en trois parties : Gestion de projet, Production et Ressources.

La cellule de pilotage de BED travaille sur Google DRIVE pour les dossiers en cours, principalement de l'année ou à cheval sur plusieurs années.

Le lien associé à cet espace est (visible en mode LECTURE):

<https://drive.google.com/drive/folders/1kMnjaOUIMZheAfQbiWVy8K0Dtp9YvBNa?usp=sharing>

Les personnes ayant accès sont décrites ci-après.

Nom	Rôle
Cécile PIGNOL	Propriétaire
Fabien ARNAUD	Lecteur
Virginie GIRARD	Editeur
Isabelle CHARPENTIER	Editeur
Mathias ROUAN	Editeur
Sylvia DAMA / Sylvie DAMY	Editeur
Wilfried HEINTZ	Editeur
Zone Atelier Alpes	Editeur

### 2b. GÉOCATALOGUE

Le géocatalogue du RZA appelé "LTER France Metadata Node" permet de visualiser les données et de structurer la recherche à l'échelle nationale et à travers le réseau (action transversale). Il moissonne les données des géocatalogues

des ZAs lorsqu'ils existent ou accueille les fiches de métadonnées produites par les ZAs (cf. partie suivante). Il est moissonné par le géocatalogue de eLTER appelé "DIP BIOSENSE".

**\*\*Géocatalogue LTER France Metadata Node**

Lien : <http://meta.data-za.org/geonetwork/srv/fre/catalog.search#/home>

Application Géocatalogue : GeoNetwork

Version GeoNetwork : v. 3.6.0.0 + cf. annexe pour les détails de la configuration

Aspects techniques / A COMPLETER

Administrateur de la machine : Mathias Rouan, Isabelle Charpentier, ??

**\*\*Géocatalogues des ZAs**

cf. Annexe - lien des géocatalogues, des mots clefs moissonnés et des personnes contact.

**\*\*Géocatalogue DIP-BIOSENSE**

Lien: <http://dip.biosense.rs/#/geoNetwork>

Ce géocatalogue moissonne le LTER French Metadata Node mais également des catalogues des partenaires eLTER EU.

Personne contact: Christoph Wohner : [christoph.wohner@umweltbundesamt.at](mailto:christoph.wohner@umweltbundesamt.at)

**\*\*Géocatalogue de INEE INDORES**

Lien : <https://www.indores.fr/geonetwork/srv/fre/catalog.search#/home>

Ce géocatalogue institutionnel du CNRS (CNRS-INEE/MNHN) moissonne le LTER French Metadata Node mais également des catalogues de laboratoire en dehors du RZA.

Personne contact: Amandine Hénon ([amandine.henon@mnhn.fr](mailto:amandine.henon@mnhn.fr)) de BBEES.

## 2c. PRODUCTION DES FICHES DE MÉTADONNÉES

*2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?*

Les producteurs de données remplissent un fichier avec des métadonnées au format de l'outil geoflow, dont le standard des métadonnées est Dublin Core.

Une routine sous R au format .json « Configuration\_rza.json », disponible sur LIEN, permet de réaliser :

- la fiche de métadonnées au format .XML à destination du geonetwork LTER French Data Node suivant la norme ISO 19115/19119 ;
- la fiche de métadonnées au format .EML à destination du PNDB si les données sont éligibles (thématique BIODIVERSITÉ, SOCIO-ÉCOSYSTÈME) ;
- de fournir les données compatibles avec un geoserver au format .WMS ou .WFS ;
- de créer un DOI associé à l'entrepôt ; pour celui de l'INEE, Data-InDoRES, cette fonctionnalité n'est pas possible.

L'identifiant unique des fiches de métadonnées pour les données relevant du rza et des actions transversales, sera indexée de la forme suivante "rza\_at\_nom-court\_code-aleatoire" ou "rza\_at\_nom-court\_aammjj" pour les actions transversales, sinon "rza\_nom-court\_code-aleatoire" ou "rza\_nom-court\_aammjj".

Le dictionnaire des données, la documentation (ex. publications, protocole, carte) ou lien vers une page web (ex. github pour les codes sources) pour chaque jeu ou base de données seront associés aux fiches de métadonnées.

Un guide d'utilisation de geoflow (LIEN) et un guide de recommandation (LIEN) sont fournis pour guider les producteurs de données dans le choix des mots clefs, des licences, etc.

## 2c. CONTRÔLE QUALITÉ ET OPÉRATION DE CURATION DES DONNÉES

*2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?*

Le contrôle de la qualité des métadonnées est réalisé sur la base d'une grille de critère par un membre de BED et d'un algorithme permettant de s'assurer de la complétude de la fiche. La fréquence des contrôles est bi-annuel.

# 3. Stockage et sauvegarde pendant le processus de recherche

## 3.a. STOCKAGE ET SAUVEGARDE DES DONNÉES

*3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?*

Le stockage et la sauvegarde des données des ZAs et des actions transversales varient selon les producteurs jusqu'à la mise en dépôt des données, sur le serveur Data-InDoRES ([cf. Annexe PGD des ZAs](#)).

## 3b. SECURITE DES DONNEES

*3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?*

Les données entreposées ne sont pas des données sensibles. Les données seront accessibles via un lien de téléchargement depuis la landing page de Data-InDoRES ([LIEN à mettre quand il sera en ligne](#)).

# 4. Exigences légales et éthiques, codes de conduite

## 4a. PROPRIETE DES DONNEES ET CONDITIONS D'ACCÈS ET D'UTILISATION

*4a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?*

Lorsque les données arrivent dans un entrepôt, elles sont déjà anonymisées et conformes au RGPD.

*4b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?*

La licence apposée sur les données est CC-BY 2.0 (<https://creativecommons.org/licenses/by/2.0/>).

Il n'existe aucune restriction d'accès sur les données; le lien d'accès aux données sera mentionné dans les fiches de métadonnées.

Le téléchargement des données via l'entrepôt de stockage des données DOIisées sera suivi à terme afin d'évaluer les ressources.

Une fois les données entreposées sur le Dataverse de l'INEE (Data-InDoRES) géré par BBEES, l'utilisation des données est contrainte à la citation du DOI, des producteurs et à la citation dans les remerciements. Chaque fichier « read me.txt » comporte les mentions exactes.

*4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?*

Non concerné.

## 5. Partage des données et conservation à long terme

### 5a. PARTAGE DES DONNÉES ET EMBARGO

*5a. Comment et quand les données seront-elles partagées ? Y-a-t-il des restrictions au partage des données ou des raisons de définir un embargo ?*

La fréquence d'actualisation des données DOIées sera de l'ordre de 1 à 3 ans suivant les jeux de données.

Aucun embargo n'est fait sur les jeux de données compilés.

Le versionning des données se fera à un pas de temps de 5 ans pour minimiser les Go entreposés.

Les données seront partagées sous licence CC, via l'accès aux entrepôts de données depuis les fiches de MD.

### 5b. SÉLECTION DES DONNÉES À CONSERVER SUR LE LONG TERME

*5b. Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?*

Les données seront entreposées et sauvegardées sur l'entrepôt de données de BBEES ( LIEN ) en premier lieu. Dans quelques rares cas, elles pourront être stockées sur les entrepôts partenaires.

Les utilisateurs prévisibles/déjà identifiés de ces données sont :

- à l'échelle nationale :
- à l'échelle européenne : le réseau eLTER

### 5c. OUTILS D'ANALYSE MIS À DISPOSITION ET RENSEIGNEMENTS

*5c. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?*

Aucun outil logiciel n'est préconisé pour exploiter les données. En revanche, la méthodologie employée est associée aux fiches de MD.

### 5d. GESTION DU DOI

*5d. Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ?*

Les services de BBEES assurent la pérennité du stockage.

Les services de BBEES disposent d'un contrat avec l'INIST pour produire leur propre DOI. Le préfixe de l'identifiant du DOI est composé de :

- du préfixe associé à INEE « 10.xxxxx »
- du préfixe associé au projet « Nom-Court »

## 6. Responsabilités et ressources en matière de gestion des données

### 6a. PERSONNES CONTACT ET DATA STEWARDSHIP



6a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ?

La liste des personnes contact, assimilées au Data Steward, est donnée dans le tableau suivant.

Tableau 1. Liste des personnes impliqués dans la gestion des données à l'échelle du RZA

Nom	ZA	Laboratoire	Statuts, disciplines, projet	Actions BED	Autres
Isabelle Charpentier	ZAEU	ICUBE	CR CNRS	Metadata curateur	
Sylvie Damy	ZAAJ	ChronoEnv	MdC informatique	Metadata curateur	
Virginie Girard	ZAA-RZA	LECA	IR Coordination, Projet Sentinelles des Alpes	Metadata curateur/producteur	INEE
Sandrine Glatron	ZAEU	DynamE	DR CNRS, Paléosciences	Metadata producteur	
Wilfried Heintz	ZAPYGAR	Dynafor	IE INRAE (Géomatique Data) CATI	Metadata curateur	INRA CATI
Françoise Le Moal	ZABrl	ECOBIO	IE (Géomatique Data)	Metadata curateur	OSUR
Emilie Lerigoleur	ZAPYGAR	GEODE	IE Géomatique, Data / LabEx DRIIHM - OHM Haut Vicdessos / Projet SO-DRIIHM	Metadata curateur	INEE
Christine Plumejeaud	ZAPVS	LIENSS	IR Data scientist		INEE, Rbdd
Cécile Pignol	ZAA	EDYTEM	IE Informations scientifiques, Projet Cyber-carothèque Nat.	Producteur Data/Ech Metadata curateur	INEE
Mathias Rouan	ZABrl	LETG	IR Data scientist (Géomatique Data)	Metadata curateur	INEE
David Sarramia	ZATU	LPC	MdC en informatique, CEBA	Metadata curateur	IN2P3
Amandine Henon	-	BBEES		Administratrice de système d'information	INEE, MNHN

Tableau 2. Lien des géocatalogues des ZAs et des personnes en charge des géocatalogues. Les ZAs soulignées ne disposent pas de leur propre Geonetwork ; (\*) ZAAJ, ZAM : Protocole OAI-PMH

Zone Atelier	Liens vers le Géocatalogue	Contact(s)
Alpes	<a href="http://leca-bdgis.u-ga.fr/geonetwork/srv/fre/catalog.search#/home">http://leca-bdgis.u-ga.fr/geonetwork/srv/fre/catalog.search#/home</a>	<a href="mailto:julien.renaud@univ-grenoble-alpes.fr">julien.renaud@univ-grenoble-alpes.fr</a> (en soutien sur la machine virtuelle OSUG <a href="mailto:eric.maldonado@inrae.fr">eric.maldonado@inrae.fr</a> )
Antarctique et Subantarctique	NA	<a href="mailto:jacques.labonne@inrae.fr">jacques.labonne@inrae.fr</a>
<u>Arc Jurassien*</u>	<a href="https://dataosu.obs-besancon.fr/">https://dataosu.obs-besancon.fr/</a>	<a href="mailto:sylvie.damy@univ-fcomte.fr">sylvie.damy@univ-fcomte.fr</a>
<u>Armorique</u>	<a href="https://www.osuris.fr/geonetwork/srv/fre/catalog.search#/home">https://www.osuris.fr/geonetwork/srv/fre/catalog.search#/home</a>	<a href="mailto:francoise.le-moal@univ-rennes1.fr">francoise.le-moal@univ-rennes1.fr</a>

Bassin du Rhône	<a href="https://metazabr.irstea.fr/geonetwork/srv/fre/catalog.search#/home">https://metazabr.irstea.fr/geonetwork/srv/fre/catalog.search#/home</a>	anne.clemens@zabr.org
Brest Iroise	<a href="http://portail.indigeo.fr/geonetwork/srv/fre/catalog.search#/home">http://portail.indigeo.fr/geonetwork/srv/fre/catalog.search#/home</a>	mathias.rouan@univ-brest.fr
<u>Environnementale Urbaine</u>	<a href="https://www.geograndest.fr/geonetwork">https://www.geograndest.fr/geonetwork</a>	A identifier par intermédiaire de Isabelle Charpentier
<u>Hwange</u>	<a href="https://sig.oreme.org/geonetwork/srv/fre/catalog.search#/home">https://sig.oreme.org/geonetwork/srv/fre/catalog.search#/home</a>	A identifier par intermédiaire de Simon Chamaillé
<u>Loire</u>	<a href="https://www.doterr.fr/geonetwork/">https://www.doterr.fr/geonetwork/</a>	A identifier par intermédiaire de Mathieu Bonnefond
<u>Moselle*</u>	<a href="https://ordar.otelo.univ-lorraine.fr/">https://ordar.otelo.univ-lorraine.fr/</a>	A identifier par intermédiaire de Marie-Noëlle Pons
Plaine et Val de Sèvre	NA	christine.plumejeaud-perreau@univ-lr.fr
Pyrénées-Garonne	<a href="http://dynids.toulouse.inra.fr:8080/geonetwork">http://dynids.toulouse.inra.fr:8080/geonetwork</a>	wilfried.heintz@inrae.fr
Seine	NA	david.eschbach@upmc.fr; elebrellec@seine-aval.fr; marie.silvestre@upmc.fr
Territoires Uranifères	NA	david.sarramia@clermont.in2p3.fr

NB. Les rôles identifiés sont : *Data producteur* (à minima centralisation des données, mise en forme), *Data curateur* (Identification des données aberrantes sur la base des fiches opérateurs), *Métadonnées producteur* (à minima remplissage du fichier XLS pour décrire les métadonnées de visualisation) et *Data Analyste* (Utilisateur des données, producteur de routine pour faciliter l'analyse et le partage des données).

La rédaction du PGD est réalisée par V. Girard en collaboration avec les membres de BED et BBEES. L'actualisation du PGD se fait annuellement après chaque AG de l'année.

## 6b. RESSOURCES

6b. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?

Le stockage des données à moyen terme sur l'entrepôt de BBEES est sans limite et repose sur une analyse volumétrique établie au préalable.

L'archivage des données sur le long terme est assuré par le CINES

La visualisation des métadonnées se fait également sur le Geonetwork de l'IR RZA ou OFV.

Les dossiers associés aux données (publications, protocoles...) sont partagés sur le site internet des ZAs ; une compilation....

Ces personnes qui ne disposent pas nécessairement des compétences en informatique et gestion de données, seront formées aux outils de mise en forme ou conversion, par l'intermédiaire de formation/atelier organisé par BED-RZA, etc.

Le temps alloué à la mise en forme des données, la production des fichiers de métadonnées et le stockage sur l'entrepôt de BBEES représente près de **XX jours** par an pour les plus impliqués.

## ANNEXES

### A1. Tableau avec les liens vers les plans de gestion des données des ZAs

**A construire**

### A2. Guide technique