



Comprendre, partager, ré-utiliser les données de biodiversité

Olivier Norvez, Thomas Milon, Sophie Pamerlon, Anne-Sophie Archambeau, Thomas Bouix, Octave Cheminée, Yvan Le Bras, Solène Robert, Chloé Vinet

► To cite this version:

Olivier Norvez, Thomas Milon, Sophie Pamerlon, Anne-Sophie Archambeau, Thomas Bouix, et al.. Comprendre, partager, ré-utiliser les données de biodiversité. Patrinat (OFB-MNHN-CNRS-IRD). 2023. mnhn-04296424

HAL Id: mnhn-04296424

<https://mnhn.hal.science/mnhn-04296424v1>

Submitted on 20 Nov 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.

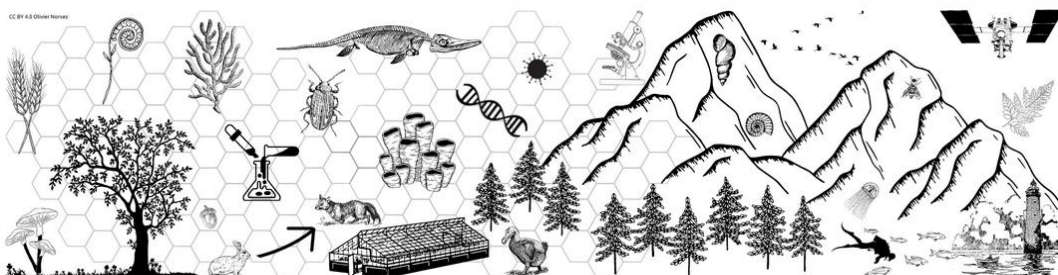


Distributed under a Creative Commons Attribution 4.0 International License

Comprendre, partager, ré-utiliser les données de biodiversité

Note explicative sur la complémentarité des systèmes d'information

Norvez O., Milon T., Pamerlon S., Archambeau A.S., Bouix T.,
Cheminée O., Le Bras Y., Robert S., Vinet C.



PATRINAT

Centre d'expertise et de données sur le patrimoine naturel

Un service commun
de l'Office français de la biodiversité,
du Muséum national d'Histoire naturelle,
du Centre national de la recherche scientifique
et de l'Institut pour la recherche et le développement



Coordinateurs du groupe de travail : Thomas Milon, Olivier Norvez, Sophie Pamerlon

Chef de l'équipe en charge du programme : Thomas Bouix

Relecteurs : Brosseau O., Damiano J., Hissel F., Lerigoleur E., Marchand F., Michal L., Pavoine S., Percevault L., Royaux C.

Référence du rapport conseillée : Norvez O., Milon T., Pamerlon S., Archambeau A.S., Bouix T., Cheminée O., Le Bras Y., Robert S., Vinet C., 2022. *Comprendre, partager et ré-utiliser les données de biodiversité : Note explicative sur la complémentarité des systèmes d'information SIB - SINP - GBIF - PNDB*. PatriNat (OFB-MNHN-CNRS-IRD) - Centre d'expertise et de données sur le patrimoine naturel. 23 pages. Etalab-2.0.

PatriNat

Centre d'expertise et de données sur le patrimoine naturel



Dans une unité scientifique associant des ingénieurs, des experts et des spécialistes de la donnée, PatriNat rapproche les compétences et les moyens de ses quatre tutelles que sont l'OFB, le MNHN, le CNRS et l'IRD.

PatriNat coordonne des programmes nationaux d'acquisition de connaissance pour cartographier les écosystèmes, les espèces et les aires protégées, surveiller les tendances de la biodiversité terrestre et marine, répertorier les zones clefs pour la conservation de la nature (ZNIEFF), et produire des référentiels scientifiques et techniques (TaxRef, HabRef, etc.). Ces programmes associent de nombreux partenaires et fédèrent les citoyens à travers des observatoires de sciences participatives (tels que Vigie-Nature, INPN espèces ou Vigie-terre).

PatriNat développe des systèmes d'information permettant de standardiser, partager, découvrir, synthétiser et archiver les données aussi bien pour les politiques publiques (SIB, SINP) que pour la recherche (PNDB) en assurant le lien avec les systèmes internationaux (GBIF, CDDA, etc.)

PatriNat apporte son expertise dans l'interprétation des données pour accompagner les acteurs et aider les décideurs à orienter leurs politiques : production d'indicateurs, notamment pour l'[Observatoire national de la biodiversité](#) (ONB) et des livrets de chiffres clés, élaboration des Listes rouges des espèces et écosystèmes menacés, revues systématiques, préparation des rapportages pour les directives européennes, élaboration d'outils de diagnostic de la biodiversité pour les acteurs des territoires, ou encore évaluation de l'efficacité des mesures de restauration. PatriNat organise également l'autorité scientifique CITES pour la France.

L'ensemble des informations (de la donnée brute à la donnée de synthèse) est rendu publique dans les portails NatureFrance, INPN et Compteur BIOM.

En savoir plus : www.patrinat.fr

Direction : Laurent PONCET et Julien TOUROULT

SOMMAIRE

Préambule.....	4
1. Comprendre le paysage, le contexte et les enjeux relatifs aux données & métadonnées de biodiversité	4
1.1. Contexte et enjeux des données et métadonnées de biodiversité.....	5
1.1.1. Cycle de la donnée et principes FAIR.....	6
1.1.2. Contexte de production des données et métadonnées.....	7
1.1.3. Données et métadonnées issues de la recherche.....	7
1.1.4. Données issues des politiques publiques.....	7
1.2. Comment déposer ses données.....	9
1.3. Descriptif des systèmes d'information.....	11
1.3.1. SIB.....	11
1.3.2. SINP.....	12
1.3.3. PNDB.....	14
1.3.4. GBIF.....	15
1.3.5. Complémentarités des SI.....	16
1.3.6. Flux de {méta}données.....	16
2. Partager les données & métadonnées sur la biodiversité	17
2.1. Pourquoi partager largement ses données ?	17
3. Utiliser les données & métadonnées sur la biodiversité	18
3.1. SIB.....	18
3.2. SINP.....	19
3.3. GBIF.....	19
3.4. PNDB.....	20
3.5. Liste des contacts et ressources associées.....	20

Préambule

Cette première version de note explicative portant sur la complémentarité des dispositifs a été élaborée dans le cadre d'un groupe de travail interne à PatriNat entre janvier 2022 et novembre 2023.

Cette note fera l'objet de mises à jour régulières et versionnées.

Les objectifs de cette note sont d'accompagner les producteurs et les ré-utilisateurs de données et métadonnées de biodiversité dans

- La **compréhension** du paysage (acteurs, qui fait quoi, qui est qui...),
- Le **partage des données/métadonnées** (où déposer, quels types de données...)
- Et l'**utilisation** de ces dernières via les complémentarités des systèmes d'informations en fonction des thématiques et/ou des publics ciblés.

Les cibles de cette note sont avant tout les chercheurs, gestionnaires, ingénieurs et techniciens données

1. Comprendre le paysage, le contexte et les enjeux relatifs aux données & métadonnées de biodiversité

L'objectif de cette partie est d'expliquer

- les enjeux relatifs aux données et métadonnées de biodiversité
- les bonnes pratiques relatives à la gestion et à l'utilisation des données de biodiversité
- le paysage des systèmes d'information de PatriNat, leurs missions, leurs objectifs, leurs complémentarités et les types d'acteurs à qui ces SI s'adressent et donc *in fine* les types de données/métadonnées que ces derniers produisent, gèrent, partagent et utilisent.

1.1. Contexte et enjeux des données et métadonnées de biodiversité

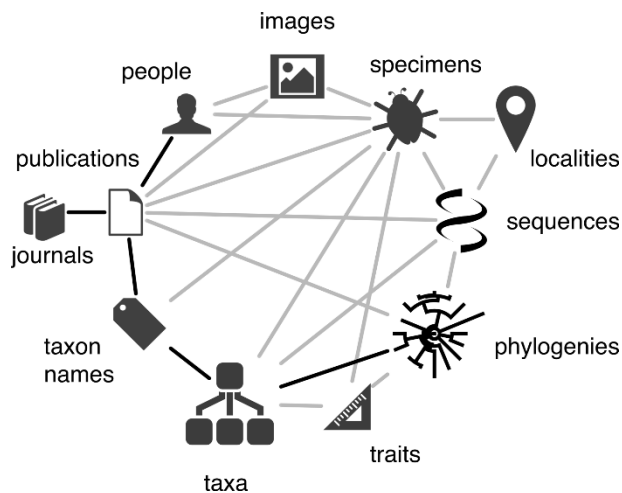


Figure 1 : Graphique de savoir sur la diversité et la complémentarité des "objets" de biodiversité et les liens entre ces derniers. Figure issue de R. Page, 2016 : doi: 10.3897/rio.2.e8767

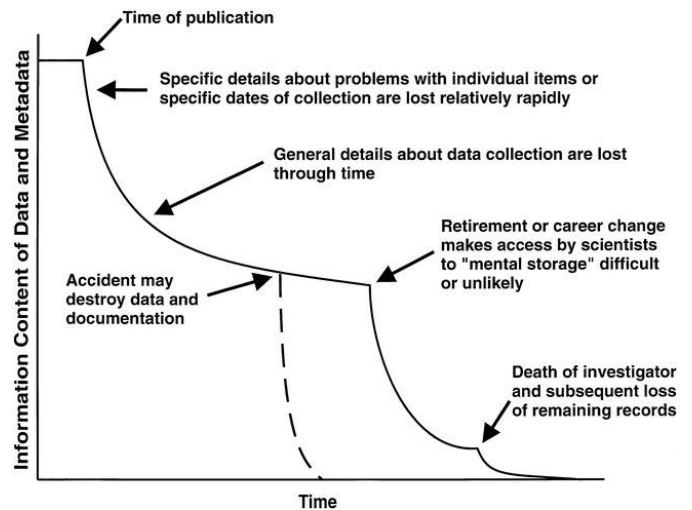


Figure 2 : Représentation de la perte d'information dans le temps à la suite d'aléas comme le manque ou la perte d'informations autour des données (métadonnées), et/ou un départ en retraite ou le décès d'un coordinateur de projet

Le contexte et les enjeux liés aux données et métadonnées de biodiversité peuvent être résumés par :

- Une certaine hétérogénéité (types de données, provenance, standards, licences, ...) & diversité "d'objets et entités" à relier entre eux¹
- Une perte d'informations dans le temps²
- Un enjeu d'ouverture et de reproductibilité³
- Des données de plus en plus massives (big data)
- Un enjeu de transversalité entre Recherche et Politiques Publiques

C'est dans ce cadre que PatriNat propose neuf grandes orientations d'actions pour accompagner et faciliter l'ouverture, le partage, et la valorisation des données de biodiversité :

- Accompagner les acteurs dans leur diversité dans le cycle d'acquisition des connaissances.
- Développer les protocoles et les outils pour la surveillance de la biodiversité des milieux continentaux et marins.
- Améliorer les processus de valorisation des données et l'interopérabilité entre les outils de gestion.
- Fournir des indicateurs robustes pour répondre aux enjeux de conservation de la biodiversité.
- Renforcer la capacité d'analyse des données au service des acteurs de la conservation de la nature.
- Proposer des méthodes et des outils scientifiques pour accompagner les entreprises et les collectivités.
- Sensibiliser la société et former les publics par l'accès aux connaissances.
- Mobiliser la connaissance disponible pour répondre à la demande sociétale.
- Renforcer les liens avec la recherche dans une approche recherche et développement.

¹ R. Page, 2016 : doi: 10.3897/rio.2.e8767

² W.K. Michener et al., 1997 doi.org/10.1890/1051-0761(1997)007[0330:NMFTES]2.0.CO;2

³ G. B. Jenkins et al., 2023 doi.org/10.1002/ece3.9961

1.1.1. Cycle de la donnée et principes FAIR

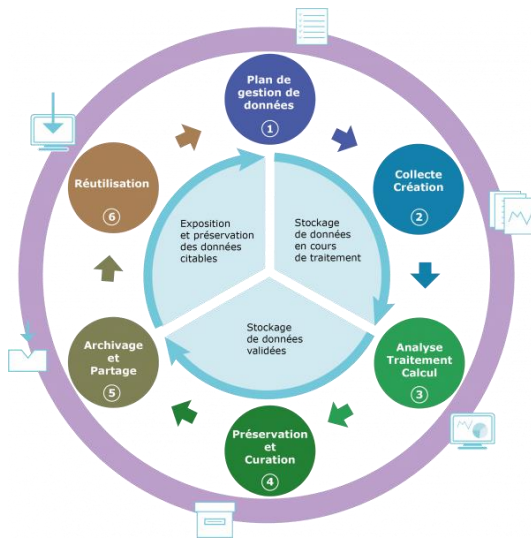
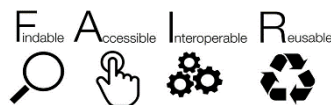


Figure 3 : schéma illustrant le cycle de vie des données.
CC-BY Nicole Lambert / GRICAD / CNRS

Le **cycle de vie des données** est l'ensemble des étapes de gestion, de conservation et de diffusion des données de recherche, associées aux activités de recherche. Pour chaque étape il existe de nombreuses ressources existantes (ex. pour la création de plan de gestion de données - PG (ou DMP en anglais)⁴. Il faut voir ce cycle de la donnée comme une boussole sur les bonnes pratiques à mettre en place au cours d'un projet.

Les **principes FAIR**⁵ sont un ensemble de principes directeurs pour gérer les données, métadonnées, codes sources, etc. et visant à les rendre **Faciles à trouver, Accessibles, Interopérables et Réutilisables** par l'humain et la machine.



Selon l'initiative internationale [GO FAIR](https://go-fair.org), ces principes directeurs se déclinent comme suit :

Faciles à trouver

- Les (méta)données se voient attribuer un identifiant global unique et persistant
- Les données sont décrites avec des métadonnées riches
- Les métadonnées incluent clairement et explicitement l'identifiant des données qu'elles décrivent
- Les (méta)données sont enregistrées ou indexées dans une ressource consultable

Accessibles

- Les (méta)données sont récupérables par leur identifiant à l'aide d'un protocole de communication standardisé
 - Le protocole est ouvert, gratuit et universellement implémentable
 - Le protocole permet une procédure d'authentification et d'autorisation, si nécessaire
- Les métadonnées sont accessibles, même lorsque les données ne sont plus disponibles

Interopérables

- Les (méta)données utilisent un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances
- Les (méta)données utilisent des vocabulaires qui suivent les principes FAIR
- Les (méta)données incluent des références qualifiées à d'autres (méta)données

Réutilisables

- Les (méta)données sont richement décrites avec une pluralité d'attributs précis et pertinents
 - Les (méta)données sont publiées avec une licence d'utilisation des données claire et accessible
 - Les (méta)données sont associées à une provenance détaillée
 - Les (méta)données répondent aux normes communautaires applicables au domaine

⁴ Cf. centre de ressources de Recherche Data Gouv : <https://recherche.data.gouv.fr/fr/page/centres-de-ressources>

⁵ 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).

1.1.2. Contexte de production des données et métadonnées

Les données de biodiversité, quelle que soit l'approche permettant de les identifier, sont produites par de nombreux acteurs et dans de nombreux contextes.

A ce jour, deux contextes principaux nous permettront d'identifier les dispositifs les plus pertinents en ce qui concerne la question de partage et d'accès aux données :

- Les données et métadonnées produites dans un cadre de Recherche
- Les données et métadonnées produites dans un cadre de Politiques publiques

1.1.3. Données et métadonnées issues de la recherche

"Les données de recherche sont définies comme des enregistrements factuels (chiffres, textes, images et sons) utilisés comme sources primaires pour la recherche scientifique, et qui sont généralement acceptés dans la communauté scientifique comme nécessaires pour valider les résultats de la recherche."

OCDE, 2007. <https://www.oecd.org/sti/inno/38500813.pdf>

"Les métadonnées, que l'on peut définir simplement comme « des données sur les données », sont un moyen de nommer les choses et de représenter les données et leurs relations."

Christine L. Borgman, 2020. <https://books.openedition.org/oep/14692>

Les données et métadonnées produites dans un cadre de recherche (personnes, laboratoires, programmes de recherche, observatoires, etc..) sont stockées et/ou valorisées via des entrepôts institutionnels et/ou thématiques et/ou des systèmes d'informations.

Ces entrepôts et SI se retrouvent notamment dans la [gouvernance du PNDB](#) via le Réseau des représentants des systèmes d'information (RSI).

Ce [RSI](#) est constitué de :

- Représentants des infrastructures du système Terre-environnement du MESR
- Représentants des entrepôts institutionnels et thématiques et des systèmes d'informations des organismes et instituts de recherche
- Représentants pour les systèmes d'informations des universités

1.1.4. Données issues des politiques publiques

A ce jour, [31 politiques publiques](#) ont été identifiées comme "faisant l'objet de production de données sur la biodiversité".

Ces données peuvent être produites par des organismes publics (Etat, établissements publics, collectivités territoriales) ou privées (entreprises, associations).

Tableau 1 : Liste des 31 politiques publiques du SIB

Accès et partage des avantages de l'utilisation des ressources génétiques et des connaissances traditionnelles	Espèces exotiques envahissantes	Paysages
Aides et redevances	Études d'impact sur l'environnement	Pêche et aquaculture en eau douce.
Aires marines protégées	Gestion adaptative des espèces	Pêche et aquaculture en mer
Artificialisation des sols	Installations classées pour la protection de l'environnement	Permis de chasser
CITES	Installations de production d'énergies renouvelables	Planification de l'eau et des milieux aquatiques
Conservation des espèces et des habitats	Inventaire du patrimoine naturel	Plans nationaux d'action
Contrôles de police de l'eau et de la nature	Inventaire national des ressources forestières	Récifs coralliens et mangroves
Directive-cadre « stratégie pour le milieu marin »	Lutte contre le changement climatique.	Réglementation des usages de l'eau
Encadrement des usages du patrimoine naturel	Mesures compensatoires	Réseaux de surveillance épidémiologique de la faune sauvage
Espaces réglementés au titre du patrimoine naturel	Natura 2000	Statistique agricole
		Trames vertes et bleues

A terme, le système d'information sur la biodiversité fédérera les données de 31 politiques publiques, chacune identifiée à travers un système d'information qui lui est propre (les « systèmes d'information métier ») dont celui sur l'inventaire du patrimoine naturel, le SINP

1.2. Comment déposer ses données

L'objectif est de ne **déposer vos données qu'une seule fois**. En effet, tous les dispositifs ont vocation à être connectés, et à partager leurs données et métadonnées selon les types d'utilisateurs ciblés dans les différents dispositifs.

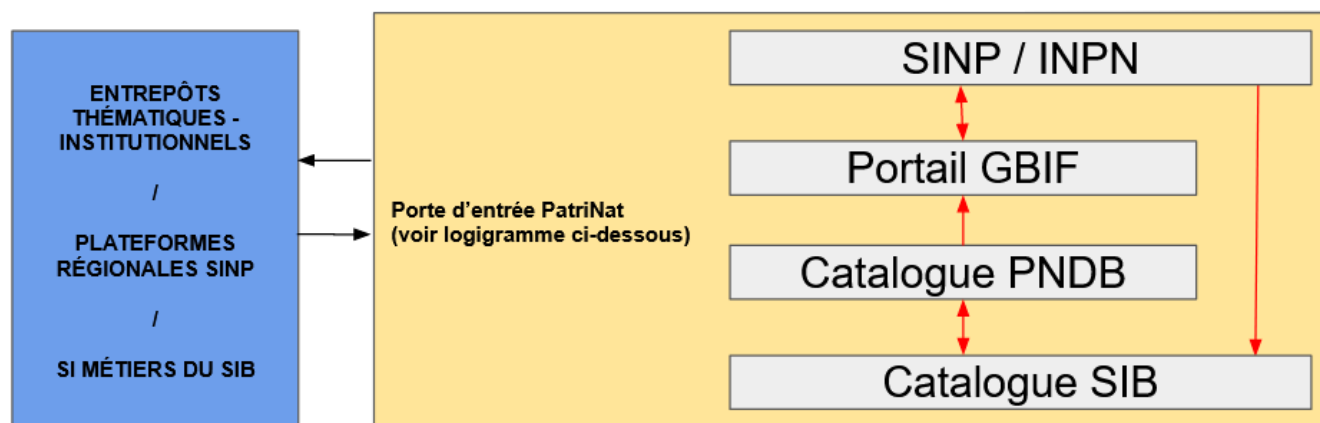
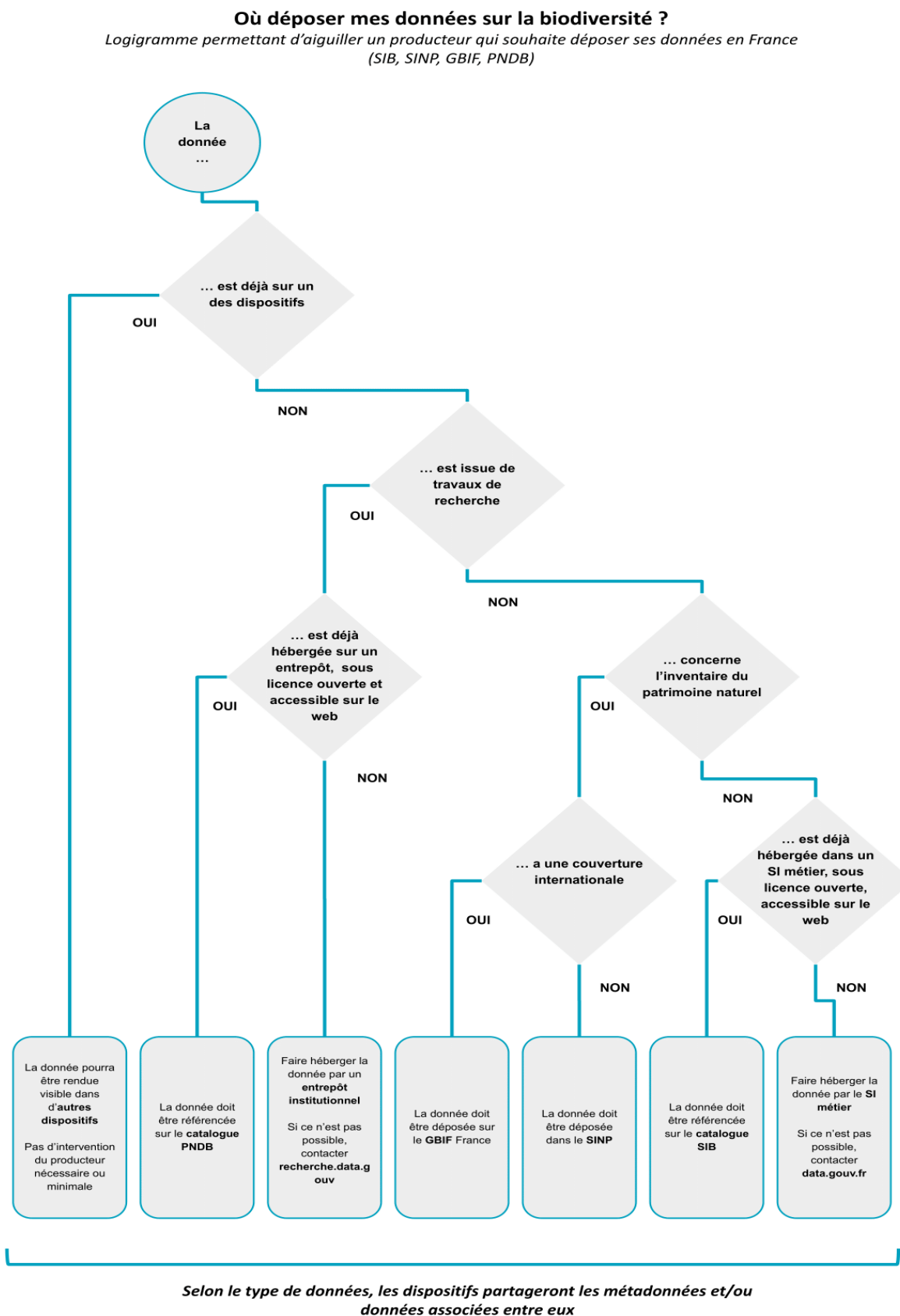


Figure 4 : schéma détaillant les liens entre dispositifs nationaux et/ou thématiques liés aux données de biodiversité

Pour définir **le point d'entrée à privilégier** pour déposer vos données, nous vous proposons de suivre le logigramme suivant (Figure 5) :

Figure 5 : logigramme permettant d'identifier le point d'entrée à privilégier pour déposer des données de biodiversité



NOTA BENE

- Dans le cadre du dispositif national [Recherche Data Gouv](#), il existe des ressources facilitant le dépôt vers le bon entrepôt et/ou l'entrepôt Recherche Data Gouv si aucune solution adaptée n'existe (entrepôts thématiques, entrepôts institutionnels, etc...). Voir le [logigramme](#).
- Les entrepôts thématiques et/ou institutionnels pour les données de recherche en biodiversité sont intégrés dans la gouvernance du PNDB via le réseau des responsables des systèmes d'informations - RSI. [Voir la liste du RSI](#)
- L'entrepôt [Easy Data \(Data Terra\)](#) offre aux communautés scientifiques **des domaines système Terre et Environnement** un service de confiance adapté aux spécificités des **données géo-référencées**. Les données de longue traîne peuvent y être déposées simplement, dans le respect des principes FAIR, et bénéficier d'un [DOI](#) pour chaque jeu de données.
- Les données et métadonnées des collections naturalistes ne sont pas concernées par cette note et doivent être déposées sur l'infrastructure dédiée à ce type de données [Recolnat](#).
- Les données et métadonnées d'étude d'impact et autres études réglementaires soumises à la procédure de dépôt légal de données brutes de biodiversité, en application des articles L411-1 A et L 122- 1 du code de l'environnement ne sont pas concernées par cette note. Ces données font en effet l'objet d'une procédure de dépôt spécifique. Cette obligation de dépôt concerne à la fois les études d'évaluation préalable à celles de suivi des impacts, réalisées dans le cadre de l'élaboration des projets d'aménagement ou des plans, schémas, programmes et autres documents de planification. Les données déposées via ce service seront prises en charge dans un second temps par le SINP. Aller sur [Depobio](#)

1.3. Descriptif des systèmes d'information

1.3.1. SIB

1.3.1.1. Pilotage / cadre légal

Créé en application de la [loi](#) du 8 août 2016 pour la reconquête de la biodiversité, de la nature et des paysages, le [SIB](#) a vocation à décloisonner les données sur la biodiversité, y compris celles d'origine privée, et à en développer les usages en proposant des services numériques adaptés.

Les détails de son organisation (périmètre, [gouvernance](#), services associés) sont définis par arrêté ministériel dans le [Schéma national des données sur la biodiversité \(SNDB\)](#). Le SNDB est le document-cadre officiel du [SIB](#).

1.3.1.2. Périmètre thématique

Le Système d'information sur la biodiversité ([SIB](#)) est un dispositif qui vise à fédérer l'ensemble des données issues de 31 politiques publiques en lien avec des enjeux sur la biodiversité, chacune identifiée à travers un système d'information qui lui est propre, appelé "Système d'information métier". (cf. Tableau 1 : Liste des 31 politiques publiques du SIB)

Le [SIB](#) cherche à faire évoluer ces systèmes vers une cohérence commune, pour rendre les données plus accessibles et réutilisables par tous. Les informations que le [SIB](#) fédère concernent tous les aspects de la biodiversité :

- son état : l'occurrence, la répartition ou encore le statut de conservation des espèces animales, végétales et fongiques, des habitats et des milieux,
- les pressions et les menaces s'exerçant sur elle,
- les actions menées en faveur de sa préservation et de sa restauration,
- les méthodes et protocoles pour la production et le partage de ces données.

1.3.1.3. Périmètre géographique

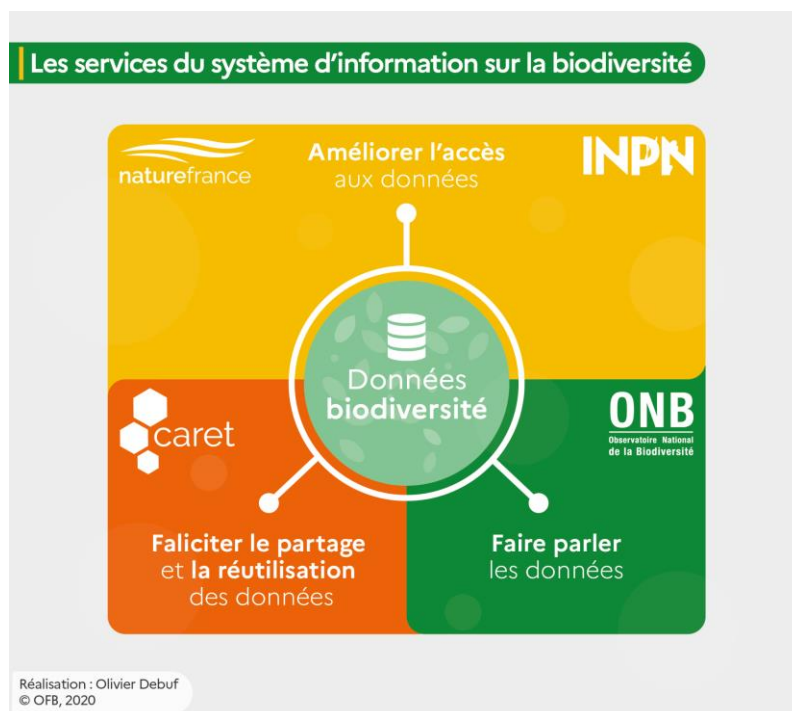
Celui des 31 politiques publiques (voir tableau p.9) qui s'appliquent sur l'ensemble ou parties du territoire français.

1.3.1.4. Nature des données gérées

Toutes les données "métiers" portant sur la biodiversité et produites par les politiques publiques, ainsi que les données de référence pour permettre l'interopérabilité, portées par le centre d'administration du référentiel technique du SIB (Caret)

1.3.1.5. Ressources pour comprendre le SI

<https://naturefrance.fr/systeme-information-biodiversite>



1.3.2. SINP

1.3.2.1. Pilotage / cadre légal

Le SINP s'inscrit dans le cadre du Plan biodiversité du 4 juillet 2018 visant à renforcer l'action de la France pour la préservation de la biodiversité et à mobiliser des leviers pour la restaurer lorsqu'elle est dégradée.

Il concourt notamment à l'axe 1 (reconquérir la biodiversité dans les territoires), l'axe 3 (protéger et restaurer la nature dans toutes ses composantes), l'axe 5 (connaître, éduquer, former) et à l'axe 6 (améliorer l'efficacité des politiques de biodiversité).

Le SINP participe également aux engagements de la Stratégie nationale pour la biodiversité (SNB), concrétisation de l'engagement français au titre de la Convention sur la diversité biologique.

Le SINP est relatif à la politique publique de l'inventaire du patrimoine naturel portée par le ministère de la Transition écologique. Le SINP est l'un des SI métiers du SIB.

1.3.2.2. Périmètre thématique

Il a pour objet de structurer les connaissances sur **l'état** et sur les **tendances** de la biodiversité au rang des espèces et des populations (faune, flore, fonge), des communautés d'espèces, des écosystèmes (les habitats naturels ou semi-naturels, ainsi que les biotopes, milieux et physionomies de végétations), et sur le patrimoine géologique, pédologique, minéralogique et paléontologique.

1.3.2.3. Périmètre géographique

Le SINP porte sur l'ensemble du territoire national terrestre, lacustre, fluvial et marin.

Il s'applique en métropole et dans les collectivités relevant du code de l'environnement (départements d'outre-mer, Terres australes et antarctiques françaises, Saint-Pierre-et-Miquelon).

Les autres collectivités ultramarines peuvent, sans que cela ne constitue une obligation, s'inscrire dans la démarche du SINP.

1.3.2.4. Nature des données gérées

- les **données d'observation et de suivi**, telles que le relevé de la présence ou de l'absence d'un taxon, d'un habitat ou d'un composant de géodiversité, les caractéristiques d'individus d'une espèce, la surface ou la structure d'un habitat ou encore la composition ou l'effectif d'une communauté d'espèces ou d'un complexe d'habitats ;
- les **données synthétiques** telles que les données issues de traitement, d'agrégation, de modélisation (variables essentielles de biodiversité, indicateurs, etc.) ou d'expertise comme par exemple les zones naturelles d'intérêt écologique, faunistique et floristique (Znieff), les sites géologiques de l'inventaire du patrimoine géologique (INPG), les cartes de répartition d'espèces ou d'habitats ou encore les effectifs et tendances temporelles par espèce ou communauté ;
- les **données descriptives**, telles que les bases de connaissance sur les traits de vie, les paramètres environnementaux ou encore les relations entre les différents composants des écosystèmes (interaction entre espèces, lien entre espèce et habitat, etc.) ;
- les **données de référence**, éléments d'interopérabilité du système d'information, telles que le référentiel taxonomique (Taxref) ou le référentiel des typologies d'habitats (Habref) ;
- les **métadonnées**, informations descriptives des données, quel que soit leur type, permettant de préciser des éléments relatifs à un lot de données (production des données brutes, production d'un référentiel, d'une synthèse, etc.).

1.3.2.5. Ressources pour comprendre le SI

<https://sinp.naturefrance.fr/> et <https://inpn.mnhn.fr/>

PNDB

1.3.2.6. Pilotage / cadre légal

Le Pôle national de données de biodiversité - PNDB - est une infrastructure de recherche (IR) virtuelle donnant accès aux données de biodiversité, de la molécule aux anthroposystèmes et sur le temps long, ainsi qu'à un certain nombre d'outils de traitement des données et autres services à la recherche.

Le PNDB joue le rôle de centre de référence thématique national pour accompagner et faciliter la compréhension, le partage, et l'utilisation des {méta}données de biodiversité. Le PNDB est le fruit d'un partenariat entre 18 établissements partenaires avec le soutien de la Fondation pour la recherche sur la biodiversité (FRB), et dont la coordination technique est assurée par le Muséum national d'Histoire naturelle via PatriNat, afin d'assurer une complémentarité d'outils et de services entre les différents systèmes d'informations sur la biodiversité (SINP/INPN, SIB, point nodal français GBIF).

Le PNDB rejoindra en 2024 l'IR rejoindra [Data Terra](#) et deviendra le cinquième pôle de données aux côtés des pôles [Odatis](#) (océan), [AERIS](#) (Atmosphère), [Theia](#) (surfaces continentales) et [Form@ter](#) (Terre solide).

1.3.2.7. Périmètre thématique

Le PNDB permet l'accompagnement et la facilitation pour la compréhension, le partage et la réutilisation des données & métadonnées de biodiversité.

Les données traitées par le PNDB concernent toutes les dimensions de la biodiversité, dans le but de prendre en compte la biodiversité sur le temps long (depuis les origines de la vie), à toutes les échelles biologiques (de la molécule à l'anthropo-écosystème), et dans toutes ses interactions.

1.3.2.8. Périmètre géographique

Toute production scientifique générée par un ou des membres d'une institution scientifique française. Donc des données de métropole, des DROM-COM (départements et régions d'outre-mer et collectivités d'outre-mer) et des données issues d'un projet de recherche international dans lequel un chercheur d'une institution française pourrait être impliqué notamment en tant qu'investigateur principal.

1.3.2.9. Nature des données gérées

Ces données proviennent essentiellement des systèmes d'informations, entrepôts institutionnels, infrastructures de recherche (observatoires, dispositifs expérimentaux, collections...), mais également des données d'expertise qui proviennent elles du SIB fédérant les données des politiques publiques. Le portail doit permettre aux communautés scientifiques françaises de partager leurs données, qu'elles concernent des échantillonnages pratiqués sur le territoire national ou à l'étranger.

Le but du PNDB est de permettre la recherche et de donner accès aux données brutes (fichiers tabulés, tableurs, média, extraction de bases de données, bases de données, etc.) utilisées par les scientifiques pour leurs analyses via un accès à un catalogue de métadonnées. Différents formats sont pris en compte (fichiers de tableurs, fichiers textes csv, etc.), fichiers SIG raster et vecteurs, fichiers média (images, sons, etc.), vues SQL de bases de données relationnelles, etc.

Enfin, toutes les données nommées entités de données dans le standard EML (ex. data entity: dataTable, spatialRaster, spatialVector, storedProcedure, view or otherEntity) utilisées en entrée et sorties d'analyses sont ainsi potentiellement concernées, mais également les logiciels d'analyse (outils/scripts), et les protocoles.

1.3.2.10. Ressources pour comprendre le SI

<https://www.pndb.fr/fr> et plus spécifiquement l'onglet "ressources" <https://www.pndb.fr/fr/ressources/>

1.3.3. GBIF

1.3.3.1. Pilotage / cadre légal

Le GBIF est un programme intergouvernemental et une infrastructure de données, créé en 2001, à l'initiative du comité scientifique de l'OCDE.

Son but principal est de promouvoir et faciliter l'accès libre et ouvert aux données sur la biodiversité.

Le réseau des points nodaux participants comprend 64 pays, 42 organisations associées, qui collaborent par le biais d'un protocole d'accord (*Memorandum of understanding*). Le réseau est coordonné par un secrétariat basé à Copenhague, au Danemark.

L'accès aux données sur la biodiversité est une nécessité scientifique et politique. Le GBIF, qui se consacre à améliorer cet accès, compte à la fois sur l'appui des nombreux autres programmes internationaux qui sont représentés au GBIF (tels que Species 2000, OBIS, etc.), et sur les points nodaux nationaux mis en place dans chacun des pays membres.

Le point nodal français a une double tâche :

- Représenter le GBIF en France, en mettant à disposition des communautés scientifiques et politiques françaises les outils et les services créés par le GBIF
- Représenter la France dans le GBIF, en s'assurant que la France soit visible et écoutée.

Le GBIF France est hébergé au Muséum national d'Histoire naturelle, sous l'égide du ministère de l'Enseignement supérieur et de la Recherche et du ministère de la Transition écologique et de la Cohésion des territoires.

1.3.3.2. Périmètre thématique

Toutes les données primaires de biodiversité, dont :

- Données génétiques (séquences, OTU)
- Occurrences (données d'observations naturalistes et spécimens en collection)
- Données protocolées (échantillonnage, suivis télémétriques, pièges photographiques et caméras...)
- Données issues de la littérature (dont celles partagées via Biodiversity Heritage Library ou Plazi)
- Données taxonomiques (listes de références, listes rouges, inventaires d'espèces, etc.)

1.3.3.3. Périmètre géographique

Monde (tous les territoires français et les données observées ou collectées sur les territoires d'autres pays par des institutions françaises)

1.3.3.4. Nature des données gérées

Données d'occurrences de taxons

Listes taxonomiques

Données d'échantillonnage

A venir : [nouveau modèle de données](#) pour prendre en compte d'autres types d'occurrences de biodiversité.

1.3.3.5. Ressources pour comprendre le SI

<https://www.gbif.org/what-is-gbif>

www.gbif.fr

1.3.4. Complémentarités des SI

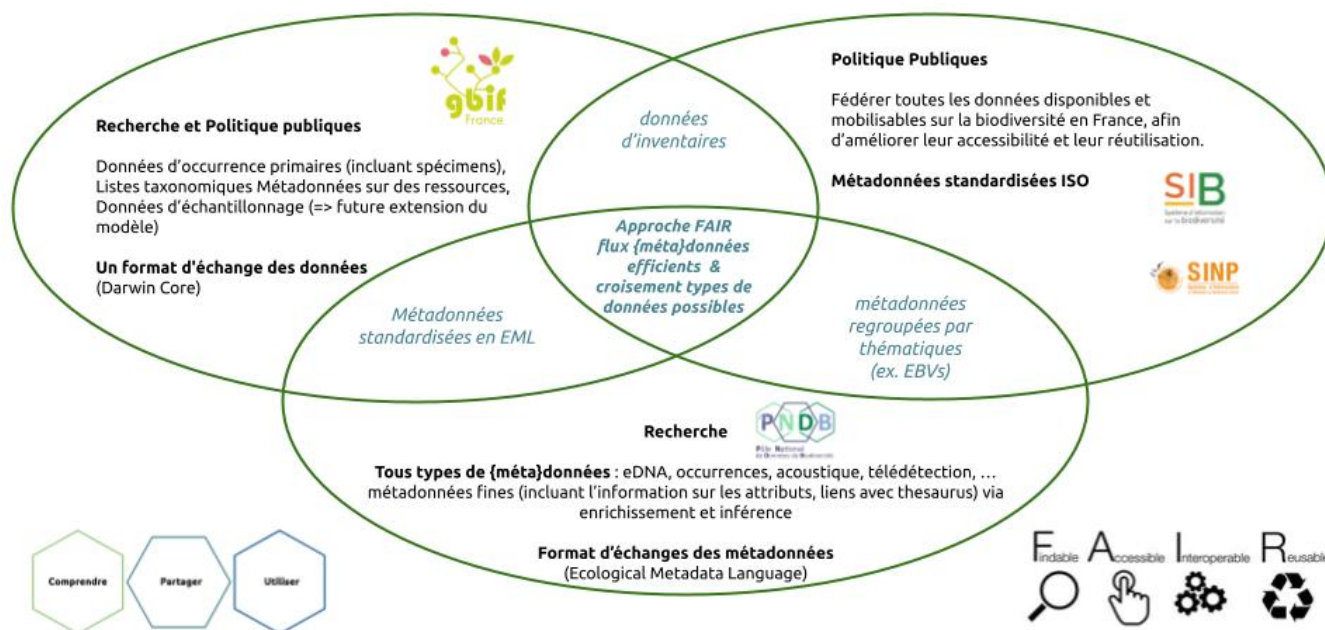


Figure 6 : complémentarités thématiques (type de données) et techniques (format de métadonnées) entre 3 systèmes d'informations agrégateurs (SIB, GBIF et le PNDB)

1.3.5. Flux de {méta}données

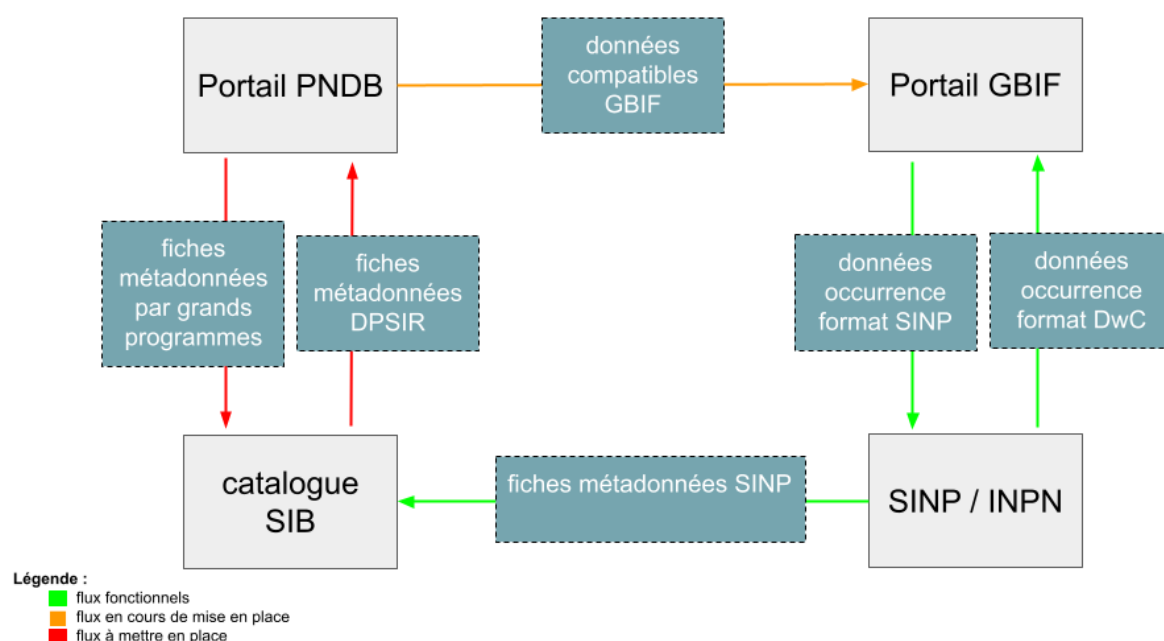


Figure 7 : schéma représentant les flux de données et/ou métadonnées entre les différents systèmes d'informations. Les {méta}données pouvant se retrouver dans tel ou tel SI le sont en fonction de leur périmètre

2. Partager les données & métadonnées sur la biodiversité

L'objectif de cette partie est de guider, quels que soient le producteur et le type de données (cartes, occurrences, ADN environnemental, etc...) dans le partage des données et métadonnées sur la biodiversité

2.1. Pourquoi partager largement ses données ?

Cette démarche globale de partage des données de biodiversité s'intègre notamment dans l'approche FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) qui est l'une des conditions *sine qua non* pour lier producteurs et ré-utilisateurs de données. Plus largement le partage des données de biodiversité permet de :

- **Valoriser et donner de la transparence à votre action** : publier en open data les données que vous produisez donne de la visibilité à votre travail et vos missions, et permet une meilleure appropriation de la vie publique en y apportant plus de transparence ;
- **Alléger votre charge de travail** : une fois le jeu de données publié, vous n'avez plus besoin de répondre à chaque demande d'accès isolée émanant d'un scientifique, d'une administration ou d'un citoyen ;
- **Contribuer à la conservation de la nature en apportant des informations factuelles** ;
- **Renforcer votre efficacité et améliorer les services aux usagers** : les données ouvertes par des administrations peuvent être réutilisées par d'autres services ou aboutir à des collaborations entre équipes, ce qui peut améliorer la mise en œuvre des missions de service public ;
- **Améliorer la visibilité et la qualité de vos données** afin de favoriser la création de nouveaux services : les données que vous mettrez à disposition seront ré-utilisées par des acteurs publics (chercheurs, gestionnaires, ...) ou privés qui pourront les croiser avec d'autres données ou détecter des anomalies. Ces données pourront être utilisées par des tiers afin de créer de nouveaux services numériques ;
- **Faciliter la réutilisation et la citation correcte** de vos données, par l'intermédiaire d'un DOI (Digital Object Identifier)⁶ associé au jeu de données, et par la publication d'un Data Paper ;
- **Éviter que d'autres refassent la même collecte de données** (ou que le décideur finance deux fois des travaux similaires) et augmenter les complémentarités entre inventaires ou programmes de recherche ;
- **Respecter le cadre légal en matière d'ouverture des données**. Le livre III du Code des relations entre le public et l'administration (CRPA) définit le cadre général de l'ouverture des données publiques. Il intègre tous les textes applicables en matière de communication, de diffusion et de réutilisation des documents administratifs. Etalab met à disposition un moteur de recherche ainsi qu'une FAQ dédiée au Code des relations entre le public et l'administration. Découvrez le CRPA simplifié ;

⁶ DOI Digital Object Identifier. Identifiant offrant un accès pérenne aux ressources numériques (publications, données, revues, rapports, etc.) grâce à un lien unique et stable. Source : <https://doranum.fr/identifiants-perennes-pid/zoom-doi/>

- **Appliquer la loi pour une République Numérique** qui vise à favoriser la "circulation des données et du savoir" à travers l'ouverture des données publiques et d'intérêt général ;
- **S'inscrire dans le cadre de la politique de « Science Ouverte »** du Ministère en charge de la recherche et de l'enseignement supérieur (MESR) à travers Le 2e Plan national pour la science ouverte, et les politiques science ouverte des établissements scientifiques et de recherche.

A noter que cette ouverture des données concerne aussi bien les données sources que les données élaborées issues d'un traitement, dans la limite des exceptions de diffusion (ex. RGPD, espèces protégées, embargo, etc.).

"Aussi ouvert que possible, aussi fermé que nécessaire" ([Rapport de la commission européenne, 2016](#))

3. Utiliser les données & métadonnées sur la biodiversité

L'objectif de cette partie est de guider, quel que soit le cadre d'utilisation, vers les bonnes ressources (catalogues, analyses, indicateurs...) adaptées aux besoins des utilisateurs de données de biodiversité (gestionnaires, chercheurs, politiques, etc.).

3.1. SIB

Le [catalogue de données](#) du Système d'[information](#) sur la biodiversité ([SIB](#)) vise à rendre accessible l'ensemble des données de biodiversité, en particulier celles des politiques publiques identifiées dans le schéma national des données sur la biodiversité, à travers les [systèmes d'information](#) métiers.

Il permet avant tout d'effectuer une recherche parmi les jeux de données déjà diffusés par ailleurs sur une plateforme de [SI](#) métier.

Ces données sont décrites dans une fiche synthétique de métadonnées (qui peut concerner plusieurs jeux de données) à l'échelle nationale et de manière intemporelle.

Le catalogue des données peut également renseigner des données qui ne sont pas encore ouvertes et documente des données de référence qui relèvent du [référentiel](#) technique élaboré par le [Caret](#) (centre d'administration du référentiel technique du SIB).

Il est progressivement alimenté par les contributions des différents [SI](#) des politiques publiques.

Ce catalogue sera directement relié au [catalogue de l'État](#), au [catalogue du PNDB](#) et aux catalogues des deux autres [SI](#) fédérateurs : [SI](#) sur le milieu marin (portail [milieumarinfrance](#)) et [SI](#) sur l'eau (portail [eaufrance](#)).

[\[Découvrir le catalogue du SIB\]](#)

3.2. SINP

Le « Requêteur National sur les données biodiversité » a pour objet la mise en place d'un outil web d'interrogation multicritères permettant la visualisation, l'exploration, la validation et le téléchargement de données d'observation d'espèces. Baptisé **OpenObs**, cet outil répond aux fonctions de la plateforme nationale du [SINP](#) en termes de diffusion de données. OpenObs présente actuellement les fonctionnalités majeures suivantes :

- Recherche simple et recherche avancée (multicritères) de données basées sur les informations taxonomiques, géographiques, temporelles et constitutives de l'observation (ex : statut de l'observation) et du jeu de données (ex : nom du jeu de données)
- Visualisation cartographique et tabulaire des résultats de la recherche
- Filtre et tri des données sur un ensemble large de critères
- Téléchargement du rendu cartographique
- Visualisation unitaire des données - Accès à l'ensemble des informations transmises par les producteurs de données
- Téléchargement des données au format SINP
- Services de redirection permettant de créer des liens depuis un site internet externe
- Interface de programmation d'application (API ou API web) qui respecte les contraintes du style d'architecture REST.

[\[découvrir Open Obs\]](#) [\[découvrir l'INPN\]](#)

3.3. GBIF

Les données disponibles sur GBIF.org sont partagées sous trois catégories correspondant aux “cores” principaux du standard Darwin Core :

- données d'occurrences d'espèces
- données taxonomiques
- données d'échantillonnage

Ces données proviennent de multiples sources (producteurs de données ou “data publishers”) et couvrent différents domaines : génomique, ADN environnemental, observations naturalistes ou par capteurs (GPS, radar, pièges photographiques ou caméras), échantillonnages, sciences participatives, spécimens en collection, données issues de culture ou élevage, données marines, études d'impact, échantillons ADN, variables environnementales et à l'échelle des communautés, fichiers multimédias, interactions biotiques, barcoding...

Elles sont accessibles via [l'interface utilisateur du site GBIF.org](#) et par [l'API GBIF](#), qui permettent de trier les données d'occurrences par l'intermédiaire de filtres spécifiques (champs taxonomiques, géographiques, etc.).

Les données sont ensuite librement téléchargeables et réutilisables selon les conditions des licences choisies par leur propriétaire ; cette réutilisation et la citation de la source des données sont facilitées par les DOI (Digital Object Identifier) attribués par le GBIF à chaque jeu de données et résultat de recherche sur GBIF.org.

Les données accessibles via le réseau GBIF sont citées dans des études scientifiques publiées dans des revues à comité de lecture, à un rythme de près de quatre articles par jour (ces articles sont ensuite compilés par thématiques (conservation des espèces, étude du changement climatique, suivi des espèces exotiques envahissantes, santé humaine et animale, agro-biodiversité...) dans la [Revue Scientifique](#) annuelle du GBIF.

L'applicabilité des données gratuites et ouvertes sur la biodiversité s'étend au-delà du milieu universitaire, car les données fournies par le GBIF sont également utilisées pour éclairer la prise de décision et les politiques publiques, ainsi que par les rapports globaux d'organismes intergouvernementaux et conventionnels à grande échelle - tels que le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) et la Plateforme intergouvernementale scientifique et politique sur la biodiversité et les services écosystémiques (IPBES).

Le GBIF développe son [modèle de données](#) pour mieux intégrer ces différents types de données.

[[découvrir le GBIF](#) et [GBIF.fr](#)]

3.4. PNDB

Pour l'axe "Utiliser", le PNDB propose :

- Un [catalogue de {méta}données](#)
 - Description fine des métadonnées
 - Technologies existantes et standard international
 - Référencement des métadonnées (provenant des SI et entrepôts biodiversité) avec un lien de téléchargement direct des données
 - Création de sous-portails disponibles (pour un projet, une thématique, ex. [EBVs](#), [IA-Biodiv Net](#))
 - Rapport d'évaluation du degré de FAIRisation des métadonnées

[[découvrir le catalogue Metacat du PNDB](#)]

- Une **plateforme d'analyse en ligne** [Galaxy-Ecology](#) permettant :
 - Nettoyage des données
 - Visualisation et calcul de tendances
 - Personnalisation et/ou utilisation de cadres analytiques robustes et reproductibles
 - Production de métriques et d'indicateurs de biodiversité

[[Découvrir Galaxy-Ecology](#)]

3.5. Liste des contacts et ressources associées

	Catalogue / ressource	Contact
SIB	Catalogue SIB / Site d'information / Liste des SI métiers	sib@mnhn.fr
SINP	Contribuer au SINP / Accéder aux données	sinp-data@mnhn.fr
GBIF	Contribuer au GBIF	connexion@gbif.fr
PNDB	Portail PNDB / liste des entrepôts institutionnels	contact.pndb@mnhn.fr

RÉSUMÉ

L'étude de l'état et de la dynamique de la biodiversité tant pour les questions de recherche que pour sa prise en compte dans les politiques publiques françaises nécessite un accès ouvert à des données de qualité le plus *FAIR* possible (Facile à trouver, Accessible, interopérable et Réutilisable).

A ce jour, plusieurs programmes coordonnés par Patrinat proposent à différentes échelles, des outils et méthodologies pour gérer, partager et analyser ces données de biodiversité. Au niveau national, ces dispositifs sont le *Pôle National de Données de Biodiversité* (PNDB), le *Système d'Information sur la Biodiversité* (SIB) (dont le *Système d'Information du Patrimoine naturel* - SINP). Au niveau international, le *système mondial d'information sur la biodiversité* - GBIF appui ces dispositifs, représenté par son point national, le *GBIF France*.

Or, le paysage français des dispositifs contribuant au partage et à la réutilisation des données de biodiversité présente *une richesse d'acteurs* (chercheurs, gestionnaires, naturalistes, citoyens, etc.), de *projets* (programmes de recherche, gestion de politiques publiques, sciences participatives) et d'*objets différents* (description de la biodiversité présente et passée, organisation et déterminants de la biodiversité, relation humain / non humain, écologie de la restauration...).

Si PNDB, SIB, SINP, GBIF concourent à l'ouverture des données, utilisent des cadres méthodologiques et promeuvent les bonnes pratiques sur tout le cycle de la donnée,

Le "paysage des données de biodiversité" n'en reste pas moins complexe, et l'objectif de cette note est d'éclairer ce paysage, de faciliter le partage des données de biodiversité et leur utilisation en fonction des complémentarités de ces systèmes d'informations et selon les thématiques ou les publics.

Patrinat (OFB-MNHN-CNRS-IRD)

Centre d'expertise et de données sur le patrimoine naturel

Jardin des Plantes

CP41 – 36 rue Geoffroy Saint-Hilaire

75005 Paris

www.patrinat.fr

