



Comprendre, partager, ré- utiliser les données de biodiversité

Note explicative sur la complémentarité des SI

Présentatrices :

Sophie Pamerlon et Alice Ainsa (PatriNat, GBIF France)

Auteurs :

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

Rencontre Afigéo

Plateformes de données numériques au service d'usages innovants

30 janvier 2024



PARTIE 1 : Comprendre les données & métadonnées : *définitions*

"Les données de recherche sont définies comme des enregistrements factuels 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>

"Une donnée d'occurrence est l'observation d'un taxon (espèce, sous espèce, ...), à une localisation, une date, avec ses observateurs, par des méthodes d'observation directes (de visu) ou indirectes (empreintes, fèces, nid, ...). Ceci recouvre les données issues d'inventaire, avec ou sans protocole, de même que les suivis temporels et toute autre étude comportant le relevé de la présence ou de l'absence d'une espèce ou d'une communauté d'espèce."

INPN, consulté le 10/01/2024

<https://inpn.mnhn.fr/programme/donnees-observations-especes/occurrences>

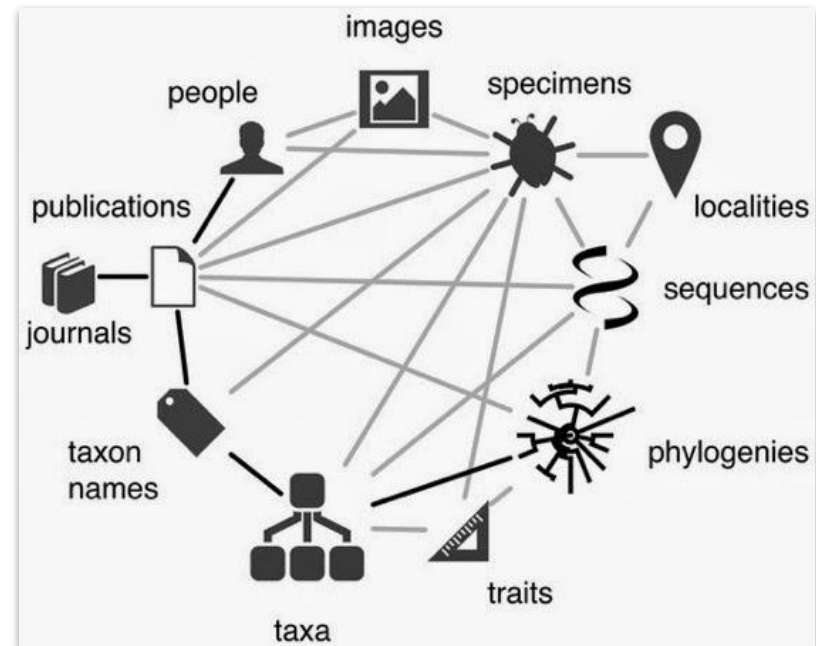
"Les métadonnées, que l'on peut définir simplement comme « des données sur les données », sont un moyen de nommer les choses et de représenter les données et leurs relations."

Christine L. Borgman, 2020.

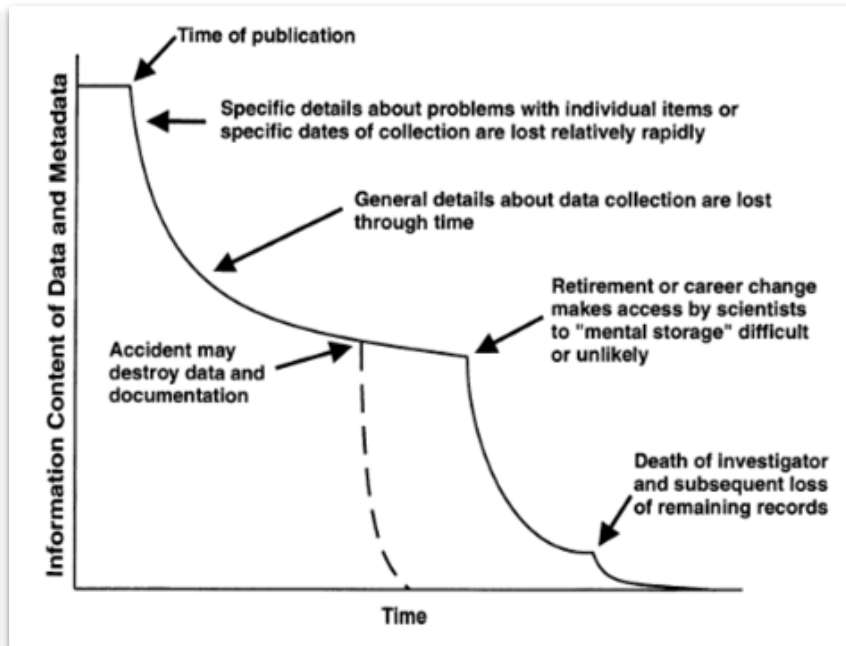
<https://books.openedition.org/oep/14692>



PARTIE 1 : Comprendre les données & métadonnées : *contexte*



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



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

Ecology and Evolution

Open Access

VIEWPOINT | Open Access | CC BY

Reproducibility in ecology and evolution: Minimum standards for data and code

Gareth B. Jenkins, Andrew P. Beckerman, Céline Bellard, Ana Benítez-López, Aaron M. Ellison, Christopher G. Foote, Andrew L. Hufton, Marcus A. Lashley, Christopher J. Lortie, Zhaoxue Ma, Allen J. Moore, Shawn R. Narum, Johan Nilsson, Bridget O'Boyle, Diogo B. Provete, Orly Razgour, Loren Rieseberg, Cynthia Riginos, Luca Santini, Benjamin Sibbett, Pedro R. Peres-Neto
... See fewer authors ^

First published: 10 May 2023 | <https://doi.org/10.1002/ece3.9961>

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

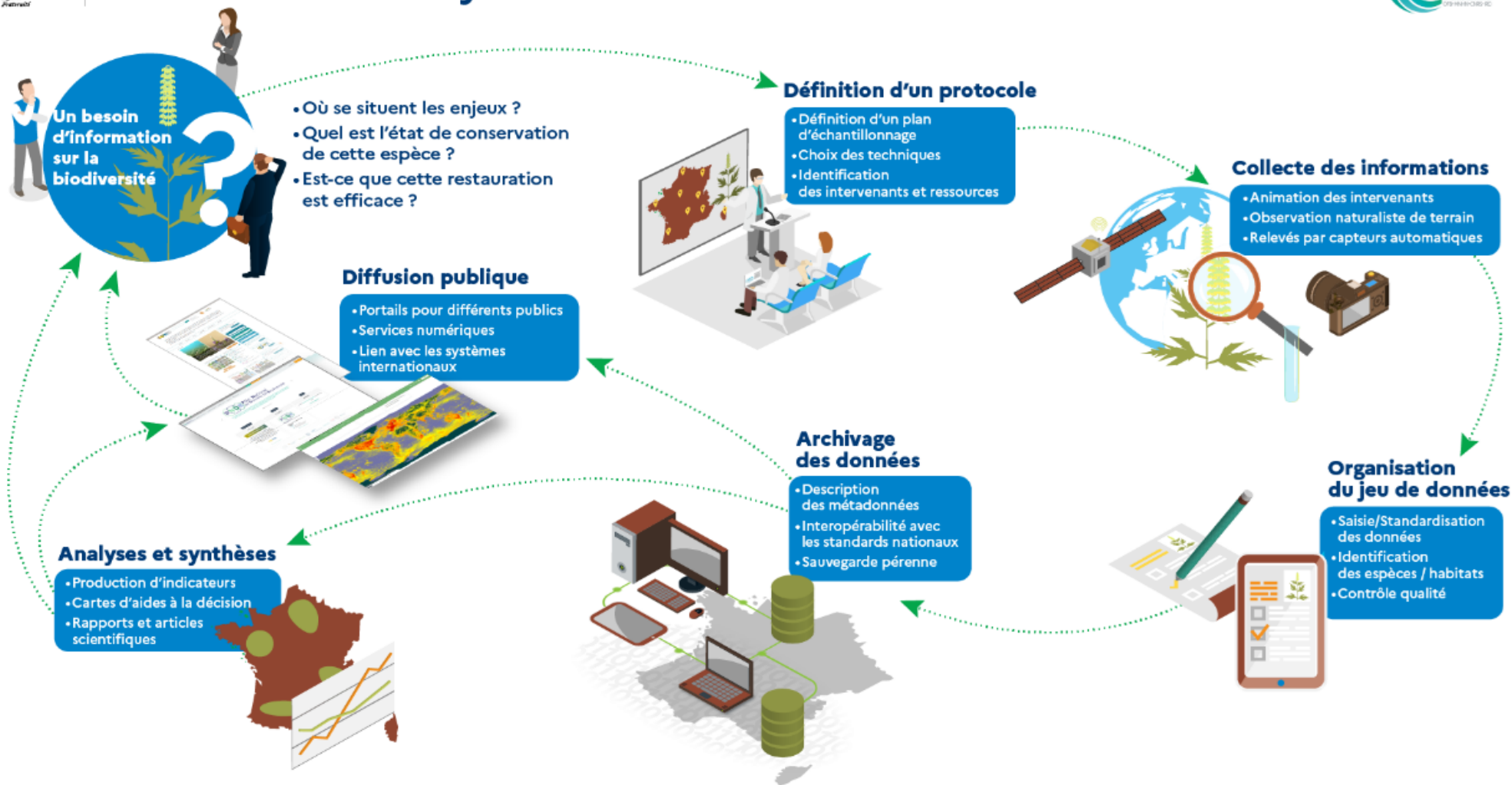
Le contexte et les enjeux liés aux données et métadonnées de biodiversité peuvent être résumés par :

- une certaine **hétérogénéité** (*types de données, provenance, standards, licences, ...*) & diversité "d'objets et entités" à relier entre eux
- une **perte d'informations** dans le temps
- un enjeu **d'ouverture et de reproductibilité**
- des données de plus en plus **massives** (=big data)
- un enjeu de **transversalité** entre Recherche et Politiques Publiques



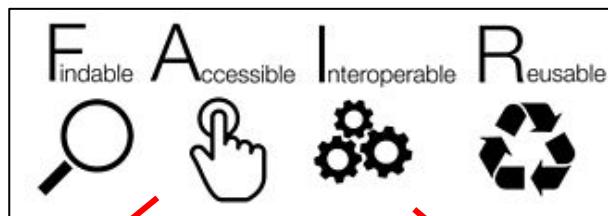
PARTIE 1 : Comprendre les données & métadonnées : *bonnes pratiques*

Cycle de la donnée biodiversité





PARTIE 1 : Comprendre les données & métadonnées : *bonnes pratiques*



Facile à trouver

Métadonnées détaillées dont renseignements des attributs

Accessible

Licence ouverte, téléchargement direct des données

Interopérable

Standards internationaux, ontologies, thésaurus, outils open source

Réutilisable

Réutilisation par mon "futur moi" et par les autres

Un ensemble de **principes directeurs** pour gérer les données de la recherche visant à les rendre **faciles à trouver, accessibles, interopérables et réutilisables** par **l'humain et la machine**.

"This is the way !"



PARTIE 2 : Constats et enjeux

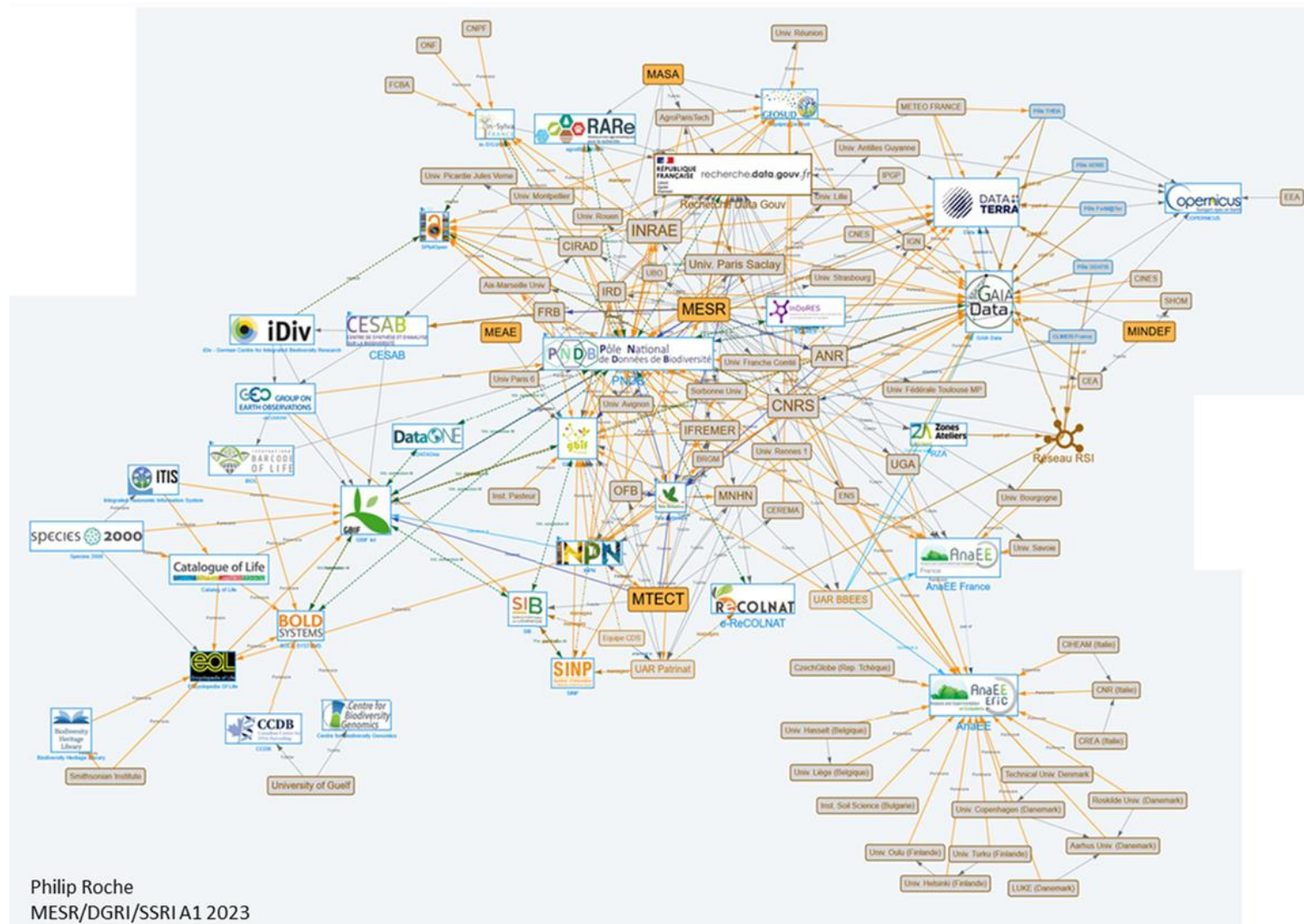
Paysage complexe des SI biodiversité :

- Diversité des outils et pratiques
- Historiques de pratiques
- Tutelles différentes
- Flux et stockage des données et métadonnées

Difficulté pour les producteurs de données à identifier le(s) bon(s) SI pour y déposer leurs données selon les scopes et thématiques

Difficulté pour les utilisateurs de savoir où et comment chercher des données

⇒ **besoin identifié d'une note de synthèse et d'un accompagnement pour mieux identifier les SI, leurs caractéristiques et leurs complémentarités**





PARTIE 3 : Complémentarités des SI biodiversité

Groupe de travail PatriNat composé du : GBIF France, PNDB, SIB, SINP/INPN (entre janvier 2022 et novembre 2023)

Les objectifs de cette note sont d'accompagner les producteurs et les réutilisateurs 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**

=> il s'agit d'une V1 !

disponible dans toutes les
bonnes librairies (HAL) !



Comprendre, partager, ré-utiliser les données de biodiversité

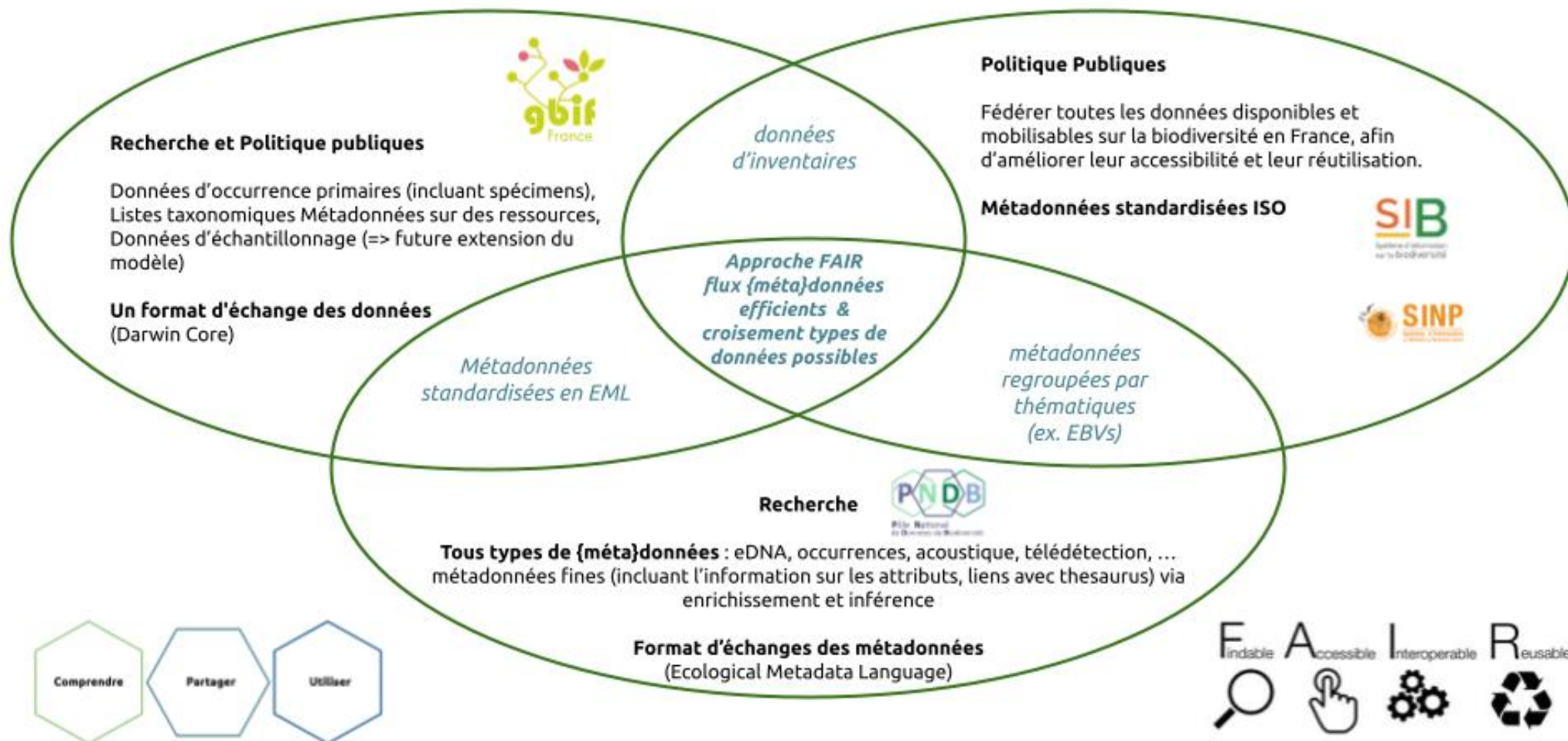
*Note explicative sur la complémentarité des systèmes
d'information*

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



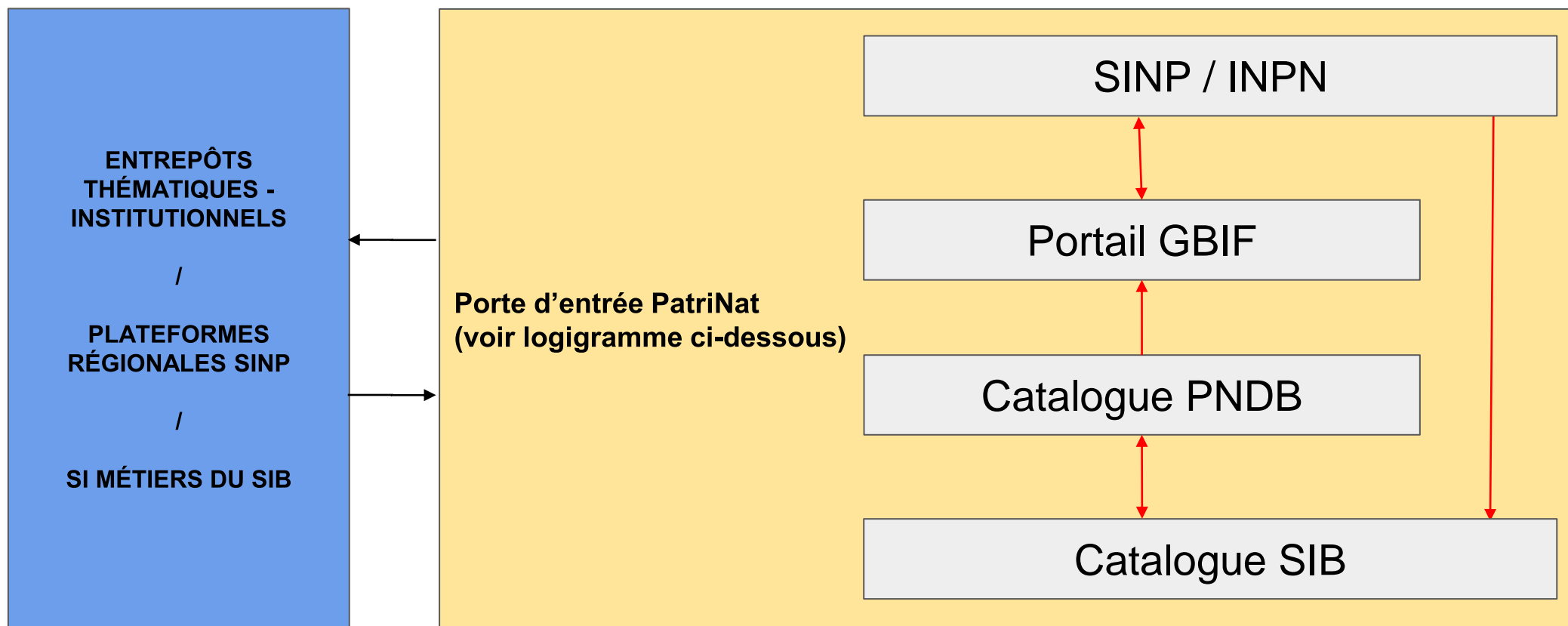


PARTIE 3 : Complémentarités des SI biodiversité





PARTIE 3 : Complémentarités des SI biodiversité





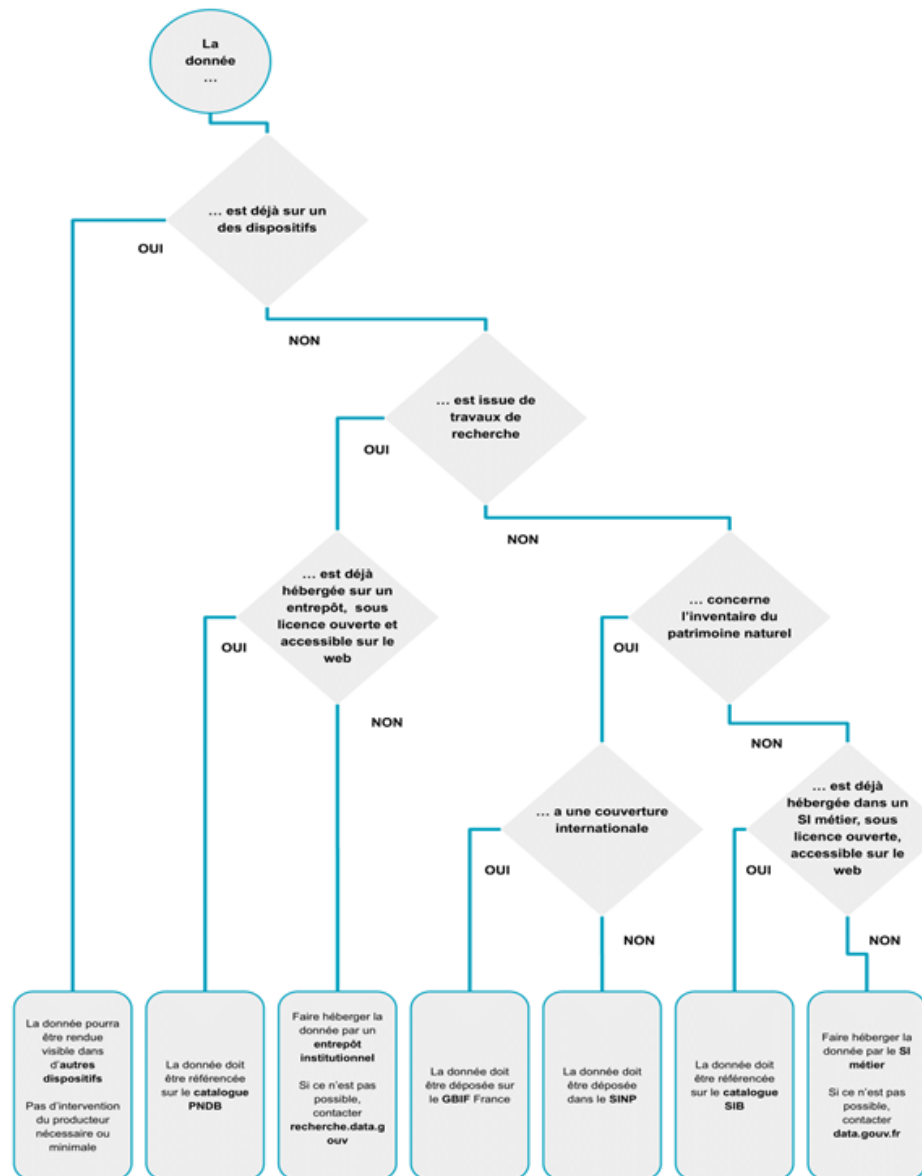
PARTIE 3 : Complémentarités des SI biodiversité



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

pages 10 et 11

Où déposer mes données sur la biodiversité ? Logigramme permettant d'aiguiller un producteur qui souhaite déposer ses données en France (SIB, SINP, GBIF, PNDB)







Selon le type de données, les dispositifs partageront les métadonnées et/ou données associées entre eux



PARTIE 4 : API et interopérabilité

Divers outils et standards ont été mis en place pour faciliter le partage, le suivi et la réutilisation des données de biodiversité

<p>Attribution</p> 	<p>Digital Object Identifiers assignés aux :</p> <ul style="list-style-type: none"> • Jeux de données publiés • Téléchargements des utilisateurs • Data papers
<p>Licences</p> 	<p>Les éditeurs choisissent une licence appropriée:</p> <ul style="list-style-type: none"> • CC-0 • CC-BY • CC-BY-NC
<p>Standards</p> 	<p>Données: DarwinCore, ABCD</p> <p>Metadonnées: Ecological Metadata Language (EML)</p>
<p>Accessibilité</p> 	<p>GBIF.org portail d'accès aux données</p> <ul style="list-style-type: none"> • Requêtes interopérables • Cartes interactives • Téléchargements gratuits • web services
<p>Publication</p>	<p>Integrated Publishing Toolkit (IPT), BioCASE, Spreadsheet templates</p>



PARTIE 4 : API

API GBIF



API Odata



API GBIF : Accès aux données de biodiversité à l'échelle mondiale.

API Odata de l'INPN : Accès aux données de biodiversité EN France avec le référentiel TaxRef.



PARTIE 4 : API GBIF

API GBIF : <https://techdocs.gbif.org/en/openapi>

Elle permet d'interroger et de publier des données sur le GBIF.org

- Intégrer des données GBIF dans d'autres sites web
- Analyse de données
- Applicable pour la recherche scientifique, surveillance de la Biodiversité, éducation, cartographie des espèces ...



PARTIE 4 : API GBIF

Documentation en plusieurs sections

GBIF API reference

Registry API

Registry API (Principal methods)

Species API

Occurrence API

Maps API

Literature API

Validator API

Vocabulary API

Analytics and Data Trends

- Registry : Pour créer éditer, mettre à jour et rechercher des informations sur l'ensemble des données.
- Species : Accès aux informations sur les espèces
- Occurrence : Accès aux informations sur les occurrences, consultation et téléchargement
- Maps : Services pour afficher des cartes contenant les informations contenues dans le GBIF
- Literature : Accès à des articles et autres documents indexés par le GBIF, citant des jeux de données et téléchargement issus du GBIF
- Validator : Permet d'exécuter des vérification sur la structure et la qualité des données avant sa publication au GBIF
- Vocabulary : Utilisé lors de l'interprétation des données afin de standardiser les valeurs pour certains champs spécifiques



PARTIE 4 : API GBIF

POST, PUT et DELETE requièrent une authentification :

```
$ curl -i --user yourUserName:yourPassword -H "Content-Type: application/json" -X POST -d @filter.json  
https://api.gbif.org/v1/occurrence/download/request`
```

OpenAPI (Open Application Programming Interface) : décrit de manière standardisée les fonctionnalités offertes par l'API

POST	/occurrence/download/request Requests the creation of a download file.	⌵
GET	/occurrence/download/request/{key} Retrieve the resulting download file	⌵
DELETE	/occurrence/download/request/{key} Cancel a running download	⌵



PARTIE 4 : API GBIF

GET `/occurrence/{gbifId}` Occurrence by id ^

Retrieve details for a single, interpreted occurrence.

The returned occurrence includes additional fields, not shown in the response below. They are verbatim fields which are not interpreted by GBIF's system, e.g. **location**. The names are the short Darwin Core Term names.

Parameters Cancel

Name	Description
gbifId * required	Integer gbifId for the occurrence.
<code>integer(\$int64)</code> <i>(path)</i>	<input type="text" value="3337906843"/>

Execute Clear



PARTIE 4 : API GBIF

Responses

Curl

```
curl -X 'GET' \
  'https://api.gbif.org/v1/occurrence/3337906843' \
  -H 'accept: application/json'
```

Request URL

```
https://api.gbif.org/v1/occurrence/3337906843
```

Server response

Code Details

200

Response body

```
{
  "basisOfRecord": "HUMAN OBSERVATION",
  "occurrenceStatus": "PRESENT",
  "taxonKey": 3089540,
  "kingdomKey": 6,
  "phylumKey": 7707728,
  "classKey": 220,
  "orderKey": 414,
  "familyKey": 3065,
  "genusKey": 3127469,
  "speciesKey": 3089540,
  "acceptedTaxonKey": 3089540,
  "scientificName": "Centaurea cheiranthifolia Willd.",
  "acceptedScientificName": "Centaurea cheiranthifolia Willd.",
  "kingdom": "Plantae",
  "phylum": "Tracheophyta",
  "order": "Asterales",
  "family": "Asteraceae",
  "genus": "Centaurea",
  "species": "Centaurea cheiranthifolia",
  "genericName": "Centaurea",
  "specificEpithet": "cheiranthifolia",
  "taxonRank": "SPECIES",
  "taxonomicStatus": "ACCEPTED",
  "iucnRedListCategory": "NE",
  "dateIdentified": "2021-08-16T14:49:40",
  "decimalLatitude": 40.592205,
  "decimalLongitude": 44.357438
}
```

[Download](#)

Response headers

```
access-control-allow-origin: *
cache-control: public,max-age=3601
content-length: 6784
content-type: application/json
expires: 0
pragma: no-cache
```

Responses

Code	Description	Links
200	Occurrence found Media type <input type="text" value="application/json"/> <small>Controls Accept header.</small> Example Value Schema <pre>{ "key": 0, "datasetKey": "3fa85f64-5717-4562-b3fc-2c963f66afa6", "publishingOrgKey": "3fa85f64-5717-4562-b3fc-2c963f66afa6", "networkKeys": ["3fa85f64-5717-4562-b3fc-2c963f66afa6"], "installationKey": "3fa85f64-5717-4562-b3fc-2c963f66afa6", "hostingOrganizationKey": "3fa85f64-5717-4562-b3fc-2c963f66afa6", "publishingCountry": "AF", "protocol": "EML", "lastCrawled": "2024-01-24T09:27:32.017Z", "lastParsed": "2024-01-24T09:27:32.017Z", "crawlId": 0, "projectId": "string", "programmeAcronym": "string", "extensions": { "additionalProp1": [{ "additionalProp1": "string", "additionalProp2": "string", "additionalProp3": "string" }], "additionalProp2": [{ "additionalProp1": "string", "additionalProp2": "string", "additionalProp3": "string" }] } }</pre>	No links
400	Invalid identifier supplied	No links
404	Occurrence not found	No links



PARTIE 4 : API Odata

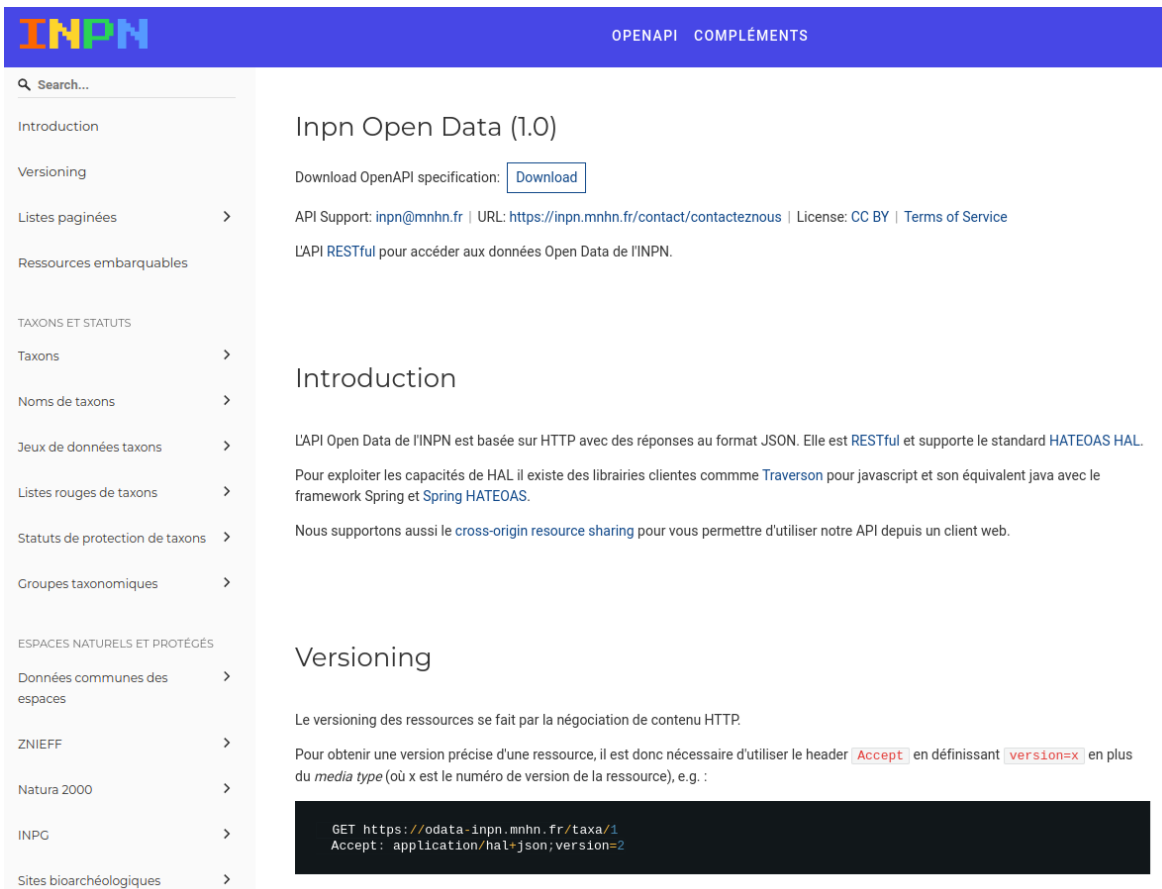
API Odata : <https://odata-inpn.mnhn.fr/docs>

Elle permet d'accéder aux données liées aux taxons de l'INPN

- Intégrer ces données dans d'autres applications web
- Analyse de données
- Applicable pour la recherche scientifique, surveillance de la Biodiversité, éducation, cartographie des espèces ...



PARTIE 4 : API Odata



INPN OPENAPI COMPLÉMENTS

🔍 Search...

- Introduction
- Versioning
- Listes paginées >
- Ressources embarquables
- TAXONS ET STATUTS
 - Taxons >
 - Noms de taxons >
 - Jeux de données taxons >
 - Listes rouges de taxons >
 - Statuts de protection de taxons >
 - Groupes taxonomiques >
- ESPACES NATURELS ET PROTÉGÉS
 - Données communes des espaces >
 - ZNIEFF >
 - Natura 2000 >
 - INPN >
 - Sites bioarchéologiques >

Inpn Open Data (1.0)

Download OpenAPI specification: [Download](#)

API Support: inpn@mnhn.fr | URL: <https://inpn.mnhn.fr/contact/contacteznous> | License: CC BY | [Terms of Service](#)

L'API RESTful pour accéder aux données Open Data de l'INPN.

Introduction

L'API Open Data de l'INPN est basée sur HTTP avec des réponses au format JSON. Elle est RESTful et supporte le standard HATEOAS HAL.

Pour exploiter les capacités de HAL il existe des librairies clientes comme Traverson pour javascript et son équivalent java avec le framework Spring et Spring HATEOAS.

Nous supportons aussi le [cross-origin resource sharing](#) pour vous permettre d'utiliser notre API depuis un client web.

Versioning

Le versioning des ressources se fait par la négociation de contenu HTTP.

Pour obtenir une version précise d'une ressource, il est donc nécessaire d'utiliser le header `Accept` en définissant `version=x` en plus du *media type* (où x est le numéro de version de la ressource), e.g. :

```
GET https://odata-inpn.mnhn.fr/taxa/1
Accept: application/hal+json;version=2
```

Utilise aussi **OpenAPI** pour la documentation

Divisé en plusieurs sections et sous sections



PARTIE 4 : API Odata

Taxons ▼

GET Lister des taxons.

GET Obtenir les données d'un nom de taxon.

GET Obtenir la taxonomie d'un taxon.

GET Lister les statuts de présence d'un taxon aux niveaux départements et secteurs marins.

GET Obtenir le groupe grand public associé à un taxon.

GET Obtenir des liens de médias associés au taxon.

GET Obtenir la listes des statuts biogéographiques d'un taxon.

GET Obtenir la description du taxon.

GET Obtenir les données d'un document source.

Noms de taxons >

Jeux de données taxons >

Listes rouges de taxons >

Obtenir les données d'un nom de taxon.

Retourne les données d'un nom de taxon par son identifiant TAXREF (correspond au CD_NOM) dans la version en cours.

PATH PARAMETERS

→ **id** required integer <int32>
Identifiant unique de la ressource.

QUERY PARAMETERS

→ **embed** Array of strings
Items Enum: "TAXONOMY" "DATASETS" "PHOTOS" "GENERAL_PUBLIC_GROUP" "OPERATIONAL_GROUPS" "ABDSM_DATA" "LAST_OBSERVATION_DATE" "BIOGEOGRAPHIC_STATUSES" "SYNONYMS"

Ressources supplémentaires embarquables :

- **TAXONOMY** : La liste des taxons définissant les rangs taxonomiques majeurs du taxon.
- **DATASETS** : Jeux de données ayant des données mailles dans lequel le taxon est présent.
- **PHOTOS** : Métadonnées des 5 photos du taxon ayant le plus haut rang INPN.
- **GENERAL_PUBLIC_GROUP** : Groupe grand public dont fait partie le taxon.
- **OPERATIONAL_GROUPS** : Groupes opérationnel dont fait partie le taxon.
- **ABDSM_DATA** : Statuts de présence du taxon aux niveaux départements et secteurs marins.
- **LAST_OBSERVATION_DATE** : Date de la dernière observation du taxon (si une localité est spécifiée dans la requête on effectue la recherche pour les observations dans cette localité).
- **BIOGEOGRAPHIC_STATUSES** : Statuts biogéographiques associé au taxon (si une localité est précisé on ne renvoie que le statuts pour la localité données).
- **DESCRIPTION** : Description du taxon en HTML.
- **SYNONYMS** : La liste des synonymes du nom valide.

GET /taxa/{id}

Serveur de production.

https://odata-inpn.mnhn.fr/taxa/{id}

Content type

application/hal+json;version=1

Copy Expand all Collapse all

```
{
  "taxrefId": 0,
  "rank": "DOMAIN",
  "domain": "CONTINENTAL",
  "statuses": "RED_LIST",
  "hasDescription": true,
  "names": {
    "binomial": "string",
    "scientific": "string",
    "scientificHtml": "string",
    "vernacular": { ... }
  }
}
```




PARTIE 4 : API Odata

https://odata-inpn.mnhn.fr/taxa/59

GET https://odata-inpn.mnhn.fr/taxa/59

Params Authorization Headers (7) Body Pre-request Script Tests Settings

Body Cookies Headers (14) Test Results

Pretty Raw Preview Visualize JSON

```

1  {
2    "taxrefId": 59,
3    "names": {
4      "binomial": "Euproctus montanus",
5      "scientific": "Euproctus montanus (Savi, 1838)",
6      "scientificHtml": "<i>Euproctus montanus</i> (Savi, 1838)",
7      "vernacular": {
8        "fr": [
9          "Euprocte de Corse (L')",
10       ],
11       "en": [
12         "Corsican Brook Newt",
13       ]
14     }
15   },
16   "rank": "SPECIES",
17   "domain": "CONTINENTAL",
18   "statuses": [
19     "RED_LIST",
20     "PROTECTED",
21     "REGULATED",
22     "EVALUATED"
23   ],
24   "hasDescription": false,
25   "_links": {

```

GET /taxa/{id}

Response samples

200

400

404

500

Content type

application/hal+json;version=1

Copy Expand all Collapse all

```

{
  "taxrefId": 0,
  "rank": "DOMAIN",
  "domain": "CONTINENTAL",
  "statuses": "RED_LIST",
  "hasDescription": true,
  - "names": {
      "binomial": "string",
      "scientific": "string",
      "scientificHtml": "string",
      + "vernacular": { ... }
    }
}

```



Comprendre, partager, ré- utiliser les données de biodiversité

Note explicative sur la complémentarité des SI

Présentateurs :

Sophie Pamerlon et Alice Ainsa (PatriNat, GBIF France)

Auteurs :

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

Rencontre Afigéo

plateformes de données numériques au service d'usages innovants

30 janvier 2024