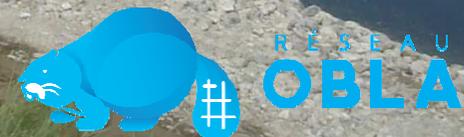


La contribution des données à un paysage international de la biodiversité :
Vers des dictionnaires d'échanges, des formats harmonisés des données.

Séverine Airaud
Ingénieur d'études pour le Réseau OBLA
(Observatoire de la Biodiversité de la Loire et de ses Affluents)
CITERES 7324 – Université de Tours



I/ L'Inventaire du Vivant:

Quoi ?

A- Gérer les connaissances acquises

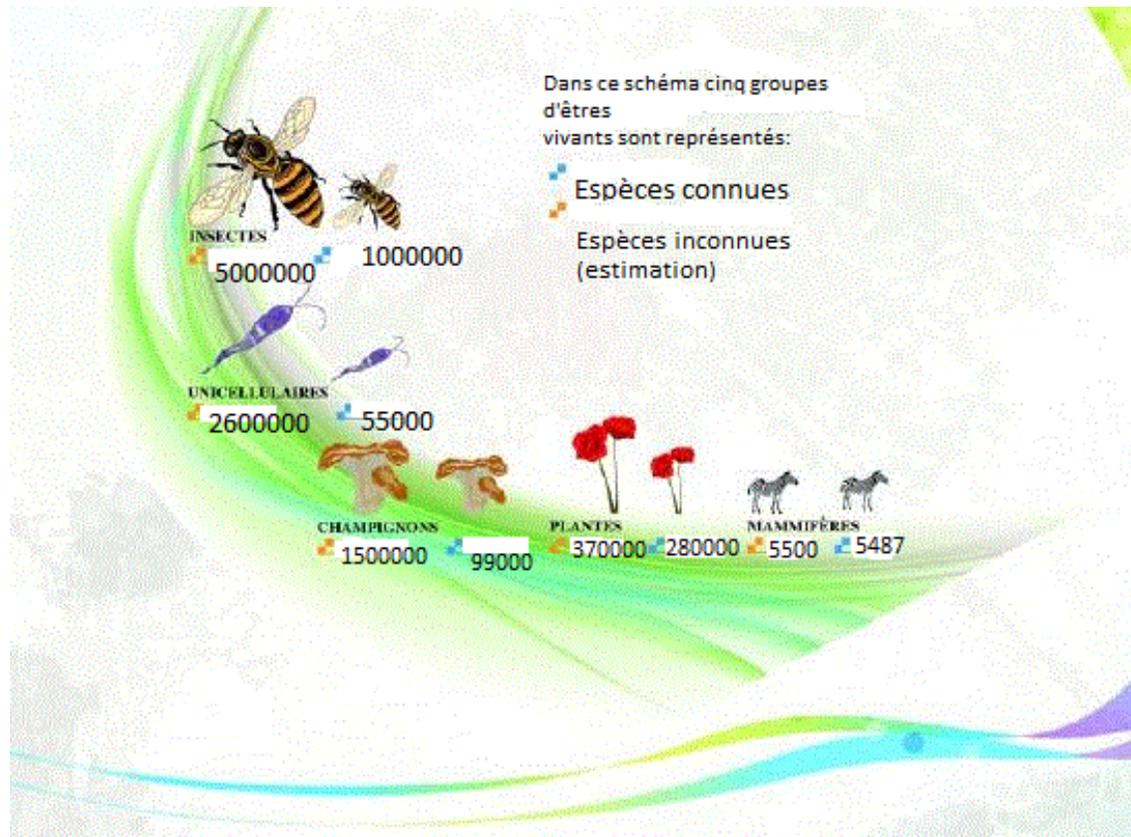


Fig.1

État des connaissances pour cinq grands groupes d'êtres vivants (représentation non proportionnelle). © A. Couty/CRDP Versailles

A- Gérer les connaissances acquises

Par ailleurs, **le stock de spécimens répertoriés dans les collections publiques est impressionnant :**

- 125 millions de spécimens pour le Muséum d'histoire naturelle américain,
- 70 millions de spécimens pour le Muséum d'histoire naturelle de Londres,
- 43 millions de spécimens pour le Muséum d'histoire naturelle à Paris,
- 7,5 millions de planches d'herbier à Kew's gardens, au Royaume-Uni,
- 11 millions de planches d'herbier au Muséum d'histoire naturelle à Paris.

B-Gérer les flux de nouvelles connaissances :

Le flux de connaissances nouvelles est important :

- chaque année, on dénombre de 16 000 à 17 000 espèces nouvelles (en majeure partie des insectes),
- en France, chaque année, des millions d'observations sont effectuées.

C- Créer des standards de données issues de la biodiversité, comment?

Modèle de données : abstraction d'un univers particulier pour le compte d'une base de données, d'un système d'information ou d'une organisation métier

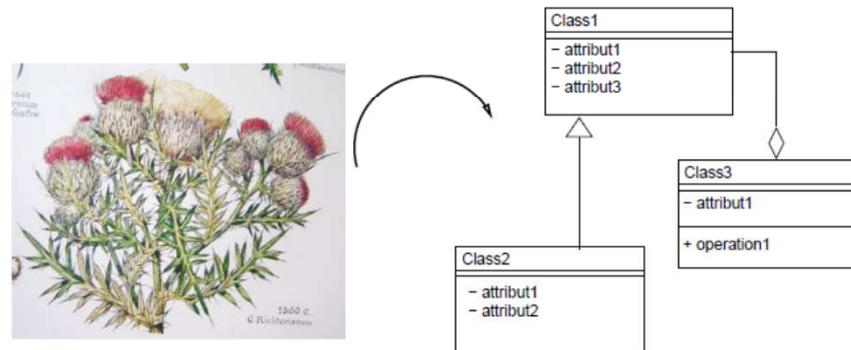


Figure: Capturer une réalité au sein d'un modèle

Fig.2 - Source : I. Mougenot

C- Créer des standards de données issues de la biodiversité, comment?

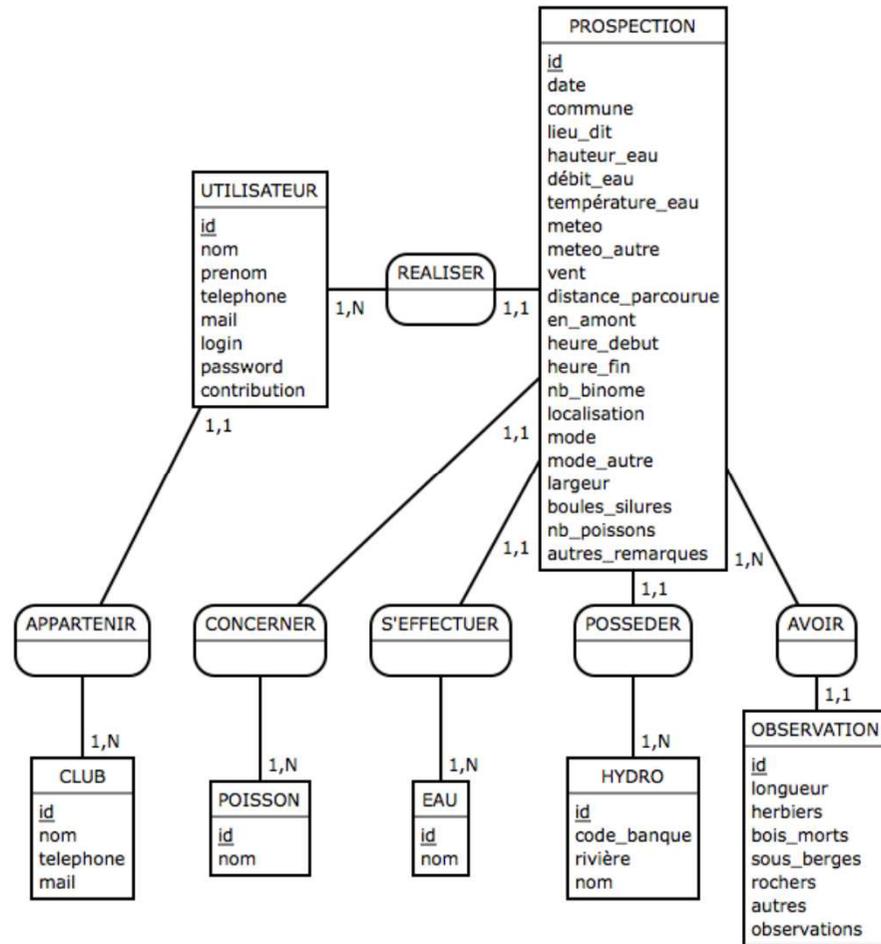


Fig.3 : MCD (Modèle Conceptuel des Données) de la BDD Silurus
(Source : A. Renaudeau pour Portail Web de la Zone Atelier Loire)

C- Créer des standards de données issues de la biodiversité, comment ?

La base de données comprend des :

Données regroupées entre elles au sein de champs et de tables. Les tables reliées entre elles forment une base de données.

Cette base de donnée comporte des entrées appelées formulaires de saisies, et des sorties : les requêtes.

Métadonnées : ce sont des données sur les données. On ajoute de l'information donnant du sens à ces données. (ex: date de la création du jeu de données, par qui, emprise géographique du jeu de données, mots-clefs).

Les métadonnées doivent être exploitables par les ordinateurs et compréhensibles par les utilisateurs.

En écologie elles doivent obéir à des normes : ISO 19139, INSPIRE... qui contribuent à leur standardisation et donc à l'interopérabilité entre les systèmes.

Elles peuvent être regroupées au sein de catalogue en ligne sur le web, tels que Geosource ou Geonetwork.

C- Créer des standards de données issues de la biodiversité, comment ?

Pour la France, deux principales structures permettent de standardiser les données issues de la biodiversité et regroupées au sein de bases de données :

- le SINP
 - le GBIF
- 

II/ Le SINP

A- Présentation de l'INPN et du SINP

Le SINP

C'est le Système d'Informations sur la Nature et les Paysages.

Il représente la plateforme de l'INPN (Institut Nationale sur la Nature et les Paysages).

L'INPN se veut une banque de connaissances sur la biodiversité et la géodiversité française.

=> INPN : plate-forme nationale

13 champs obligatoire (quoi, où, quand, qui, statut...) et env. 30 champs facultatifs (observateur, coordonnées précises...).

Le SINP – Nouveau protocole

Une approche open-data sauf données « sensibles »

•DEE : « équivalent » de la donnée source mais données standardisées, libres et gratuites

Producteur: donnée-source=> Production et gestion des données sources

Niveau régional ou thématique:=> Identifier, qualifier et diffuser les DEE régionales

National : ●Publier les référentiels

●Qualifier les données de référence

●Diffuser les DEE : recherche, visualisation, téléchargement



A- Présentation de l'INPN et du SINP

L'INPN organise, gère et diffuse l'information de nombreux programmes nationaux qui ont des méthodologies, des logiques de validation et de diffusion propres

⇒ 4 grands ensembles:

Programme « Référentiels et statuts »

- le référentiel taxonomique national « TAXREF » (et les réglementations associées)
- les référentiels « habitats »
- les référentiels géographiques...

Programme « Espaces »

- ZNIEFF, Natura 2000 et Espaces protégés français
- Inventaire national du Patrimoine Géologique (InPG)...

Programme « évaluation»

- Les listes Rouges
- Evaluation de l'état de conservation DHFF...

Programme « Espèces »

- Plate-forme nationale du SINP
- Répartitions de référence dans le cadre des inventaires nationaux
- => interactions fortes avec le GBIF France

B- Reverser ses données au SINP



Le naturaliste qui veut reverser ses données dans le système de données du SINP peut:

- Remplir directement ses données sur le site du Pôle thématique (www.naturefrance.fr).
- Transmettre ses données à la tête de Réseau concernée ou à la DREAL.
- Remplir les tables et les champs conçues par le SINP (Données Sources) , en respectant un certains nombres de champs prédéfinis et de règles (par exemple : se servir du référentiel taxonomique TAXREF).

C- Un exemple de champs à remplir pour le SINP

Relevé Faune, exemple de champs à remplir pour le SINP :

id	thematique	nom_franc	nom_latin	cd_sp	ref_sp	eff_faune	eff_precis	stade	sexe	contact
	Chiroptères					+/- 5		Adulte	Inconnu	Vu



D- De la donnée source à la DEE (Donnée Élémentaire d'Échange)

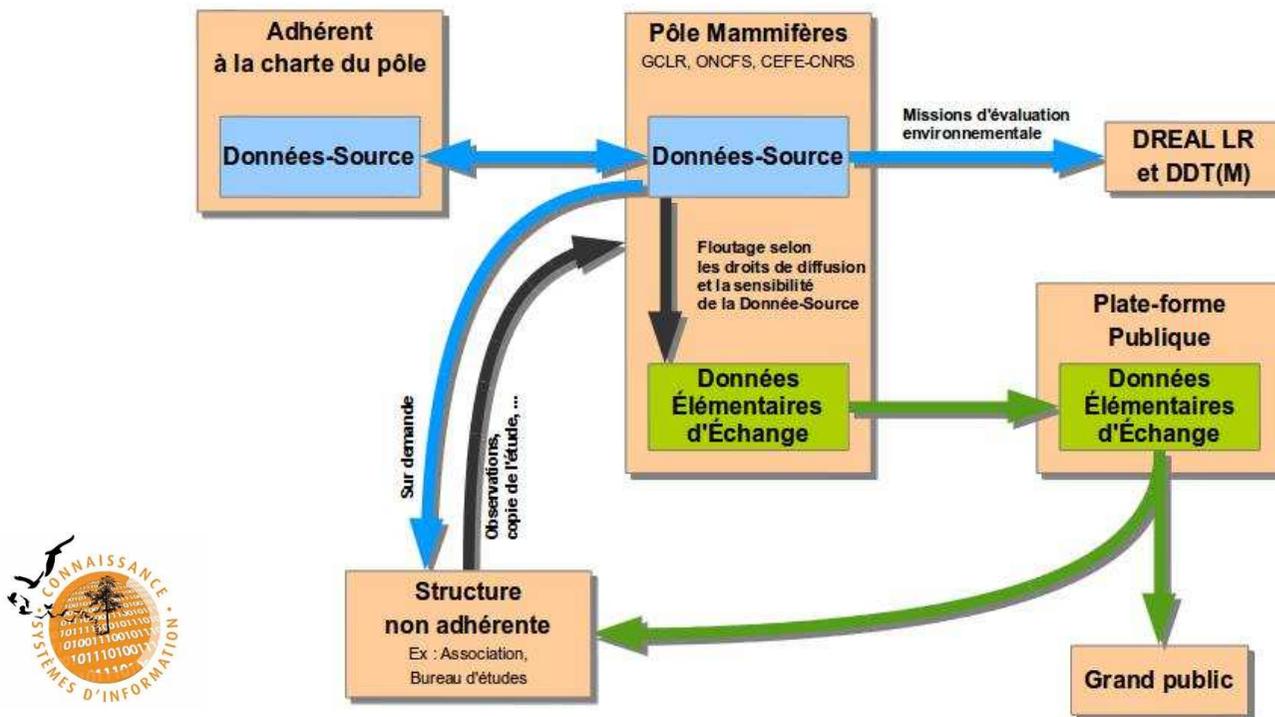


Fig.4 : de la donnée source à la donnée élémentaire d'échange (DEE)

E- Quels sont les principaux standards de données et métadonnées pour les données naturalistes ?

Données :

ABCD (TDWG Standard)

- > 1200 concepts
- XML
- Utilisé par BioCase, Tapir

Darwin Core (pre-standard v. 1.2, 47 versions)

- 48 concepts, spécimens
- XML
- Utilisé par DiGIR

Darwin Core (pré-standard v. 1.4)

- 46 concepts (plus extensions), specimens
- XML
- Utilisé par Tapir

Darwin Core (TDWG Standard)

- 172 concepts (156 dans Simple Darwin Core), données de biodiversité
- CSV, XML, RDF, JSON, ...
- Utilisé par Tapir, Darwin Core Archive ou fichier texte

Métadonnées :

EML (Ecological Metadata Language) :

EML est le second standard utilisé par le GBIF, il est utilisé pour les métadonnées. Le standard EML est largement utilisé par de nombreuses plateformes de données naturalistes comme le LTER (Long Term Ecological Research Network). EML and DwC sont les standards utilisés par les portails de données du GBIF et le GBIF Integrated Publishing Toolkit (IPT).

III/ LE GBIF (Global Biodiversity Information Facility) Système Mondial d'Information sur la Biodiversité



Le GBIF se consacre à améliorer l'accès aux données sur la biodiversité.

Le GBIF France a vocation de rassembler toutes les données hébergées en France, que celles-ci concernent la biodiversité du territoire français ou du reste du monde.

Des point nodaux ayant pour mission l'accès aux services et outils du GBIF, sont mis en place dans chacun des pays membres du GBIF.

A- Présentation du GBIF (Global Biodiversity Information Facility) Système Mondial d'Information sur la Biodiversité

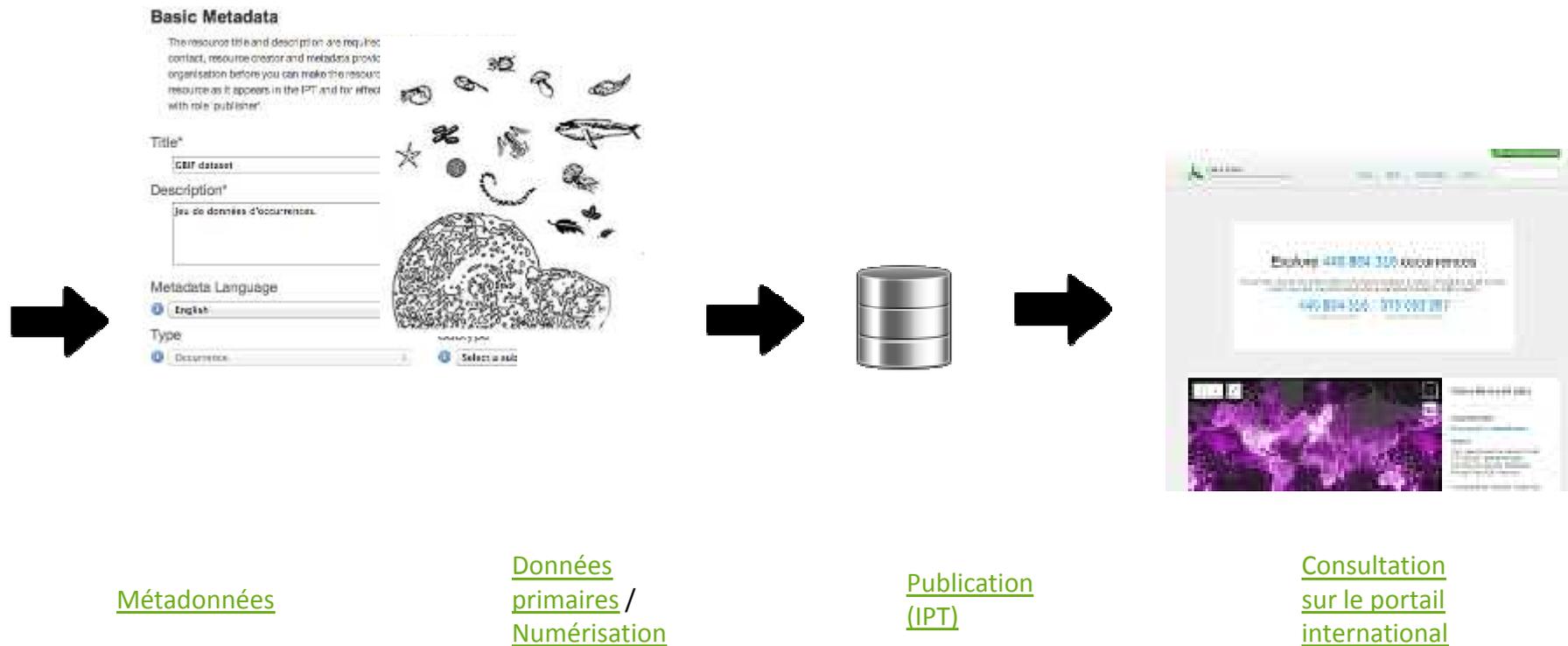


Fig. Processus de publication des données sur le portail du GBIF (source :

B- Connecter ses données au GBIF (Global Biodiversity Information Facility)
Système Mondial d'Information sur la Biodiversité



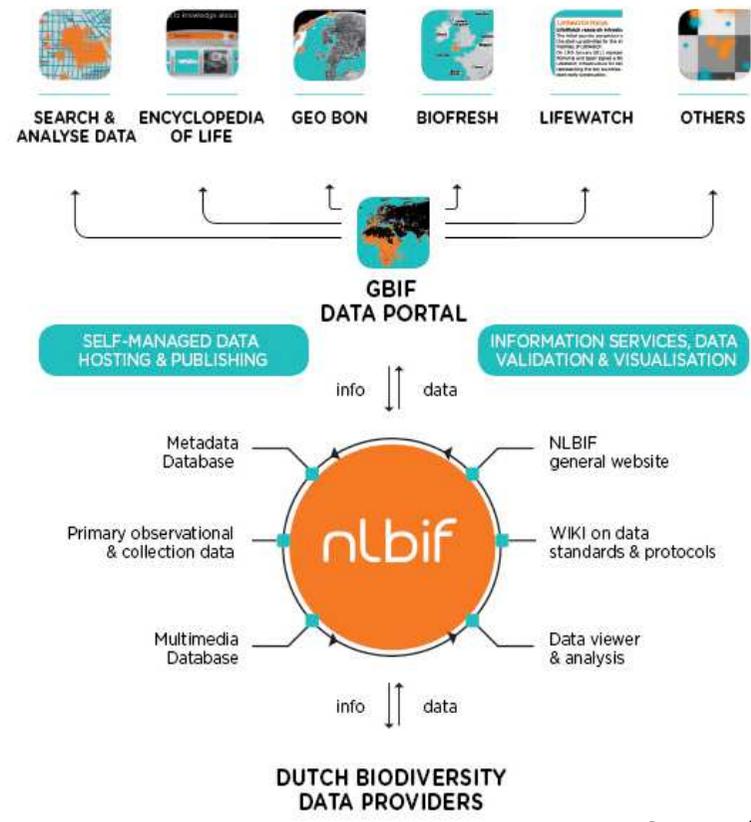
Pour connecter ses données au portail du GBIF, on utilise un logiciel « wrapper », qui permet le mapping des champs.
Pour le GBIF c'est l'IPT, il existe aussi d'autres protocoles de mapping comme

- TapirLink
- Biocase
- DIGIR

Les données sont ensuite publiées sous un format standard :

- DwC (Darwin Core), pour les données
- EML (Ecological Metadata Language) pour les métadonnées

C-Exemple de diffusion des données du Netherlands Biodiversity Information Facility (NLBIF) :



Source : <http://www.nlbif.nl/nl>

D-Vers une passerelle SINP/GBIF :

–Standard de données du SINP

- Groupe de travail pour définir un standard national

- Lien avec le standard Darwin Core (TDWG et GBIF)

- Deux objectifs différents :

- Standard « très formalisé » pour les données françaises et périmètre du SINP

- Standard international plus « ouvert »

=> Deux standards pour des besoins différents mais interopérables

- Vers un identifiant unique partagé

- TAXREF (référentiel national)

Conclusion : Conseils, Objectifs et Perspectives

La base de données naturaliste doit pouvoir être interopérable avec les bases de données nationales, comme le SINP, et internationales, comme le GBIF et donc avoir tenu-compte, lors de sa construction des « standards » de données en vigueur dans ces divers organismes :

- champs obligatoires
- logiciels utilisés pour sa construction

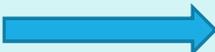
Quelques questions pourront aider dans la démarche de rendre les données Interopérables :

- Est-ce que les données sont numérisées, géoréférencées ?
- Les données sont-elles stockées dans une base de données structurée ?
- Est-ce que les données sont standardisées ?
- Les données peuvent-elles être publiées ?

Conclusion : Conseils, Objectifs et Perspectives

Pourquoi créer des standards permettant la diffusion des données issues de la biodiversité de façon internationale ?

-Pour porter à connaissance de tous et donc agir en conséquence.

-  Protéger et gérer la biodiversité

-Du point de vue du chercheur faire connaître ses travaux de recherche (possibilité de publier les données versées au GBIF sous forme de data paper).