

# PUBLICATION DE DONNÉES DE BIODIVERSITÉ AU GBIF

---

## ATELIER DE NOUAKCHOTT, MAURITANIE

2-3 SEPTEMBRE 2014



André Heughebaert  
Belgian Biodiversity Platform

# CONTENU

---

1. Les Principes du GBIF
2. Pourquoi publier ses données?
3. Rôles des intervenants
4. Propriété intellectuelle
5. Que peut-on publier?
6. Les données sensibles
7. Préparer la publication
8. Comment publier?
9. Accès aux données
10. Utilisations

# 1. PRINCIPES DU GBIF

---

- Rendre accessible l'ensemble des données primaires
- L'accès est rapide, libre et gratuit
- L'éditeur de données décide lui-même de ce qu'il publie et des éventuelles restrictions d'utilisation.
- Les droits de propriété intellectuelle ne sont en rien modifiés par la publication au GBIF.
- L'éditeur est seul garant de la qualité et de la pérénnité des données.

voir [GBIF and Open Access](#)

## 1. PRINCIPES DU GBIF

# ENGAGEMENTS DES DIFFÉRENTS ACTEURS

---

- Les pays signataires du [Protocol d'accord](#) (MoU) s'engagent à mettre en oeuvre un noeud national afin de développer le réseau dans le pays en y approuvant des institutions désireuses de publier des données.
- Les éditeurs signent un [accord de partage](#).
- Les utilisateurs signent un [accord d'utilisation](#).

# 1. PRINCIPES DU GBIF

## ACCORD DE PARTAGE

---

1. Biodiversity data accessible via the GBIF network are openly and universally available to all users within the framework of the GBIF Data Use Agreement and with the terms and conditions that the Data Publisher has identified in its metadata.
2. GBIF does not assert any intellectual property rights in the data that is made available through its network.
3. The Data Publisher warrants that they have made the necessary agreements with the original owners of the data that it can make the data available through GBIF network.
4. The Data Publisher makes reasonable efforts to ensure that the data they serve are accurate.
5. Responsibility regarding the restriction of access to sensitive data resides with the Data Publisher.
6. The Data Publisher includes stable and unique identifiers in their data so that the owner of the data is known and for other necessary purposes.
7. GBIF Secretariat may cache a copy and serve full or partial data further to other users together with the terms and conditions for use set by the Data Publisher. Queries of such data through the GBIF Secretariat are reported to the Data Publisher.
8. Data Publishers are endorsed by a GBIF Participant, if applicable, before their metadata is made available by the GBIF Secretariat.
9. GBIF Secretariat is not liable or responsible, nor are its employees or contractors, for the data contents or their use; or for any loss, damage, claim, cost or expense however it may arise, from an inability to use the GBIF network.

# 1. PRINCIPES DU GBIF

## ACCORD D'UTILISATION

---

1. The quality and completeness of data cannot be guaranteed. Users employ these data at their own risk.
2. Users shall respect restrictions of access to sensitive data.
3. In order to make attribution of use for owners of the data possible, the identifier of ownership of data must be retained with every data record.
4. Users must publicly acknowledge, in conjunction with the use of the data, the Data Publishers whose biodiversity data they have used. Data Publishers may require additional attribution of specific collections within their institution.
5. Users must comply with additional terms and conditions of use set by the Data Publisher. Where these exist they will be available through the metadata associated with the data.

# 1. PRINCIPES DU GBIF

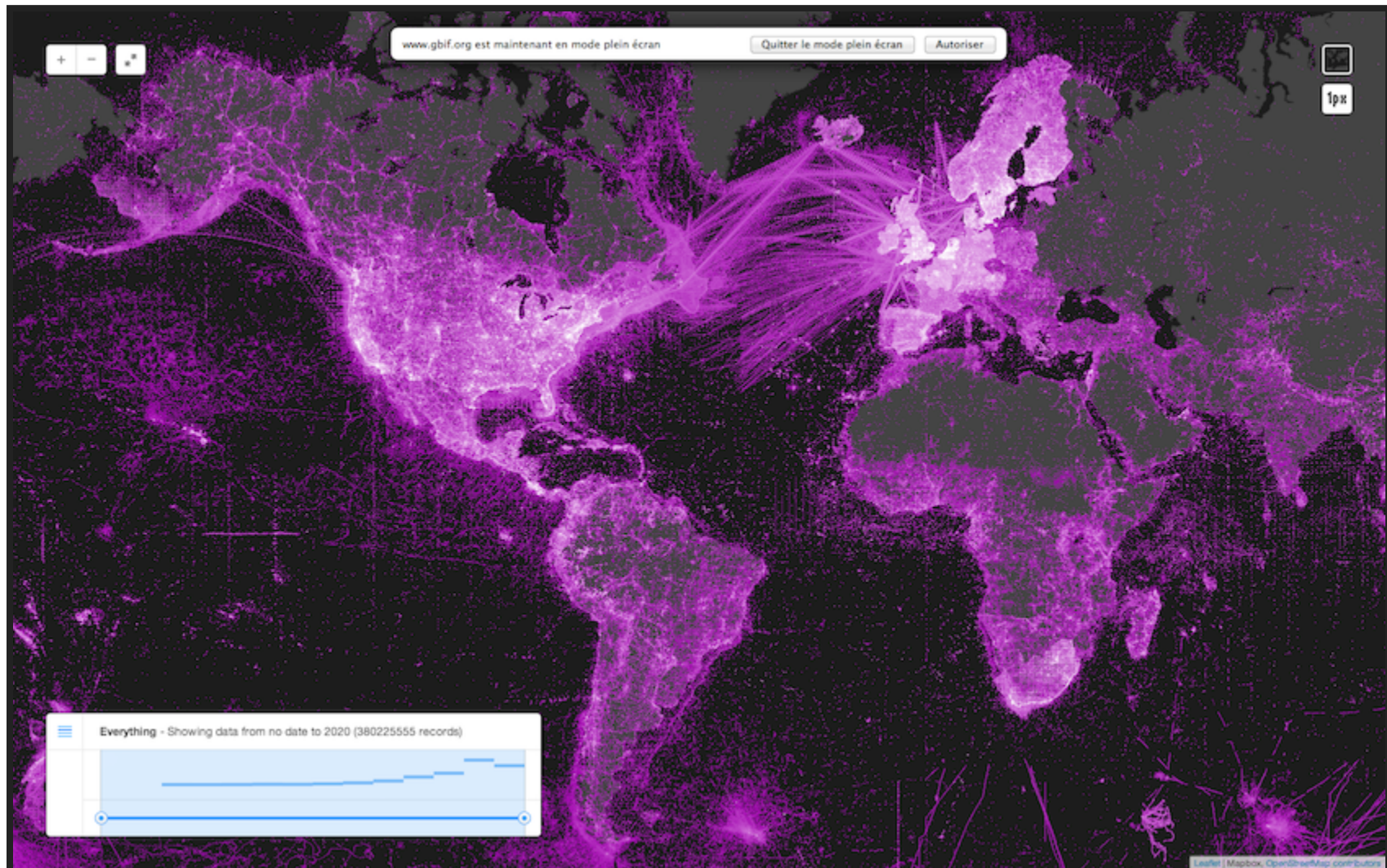
## EVOLUTION PROCHAINE

---

- La situation actuelle est assez complexe et difficile à appliquer.
- Les éditeurs seront amenés à choisir une license de type [Creative Commons](#).
- Le Secrétariat prépare, avec la communauté, un code de conduite pour les éditeurs et les utilisateurs.
- Un mécanisme de citation via le système [Digital Object Identifier](#)(DOI) est à l'étude.

voir [Consultations sur les Licenses et d'adoption des éditeurs](#)

# 2. POURQUOI PUBLIER SES DONNÉES?



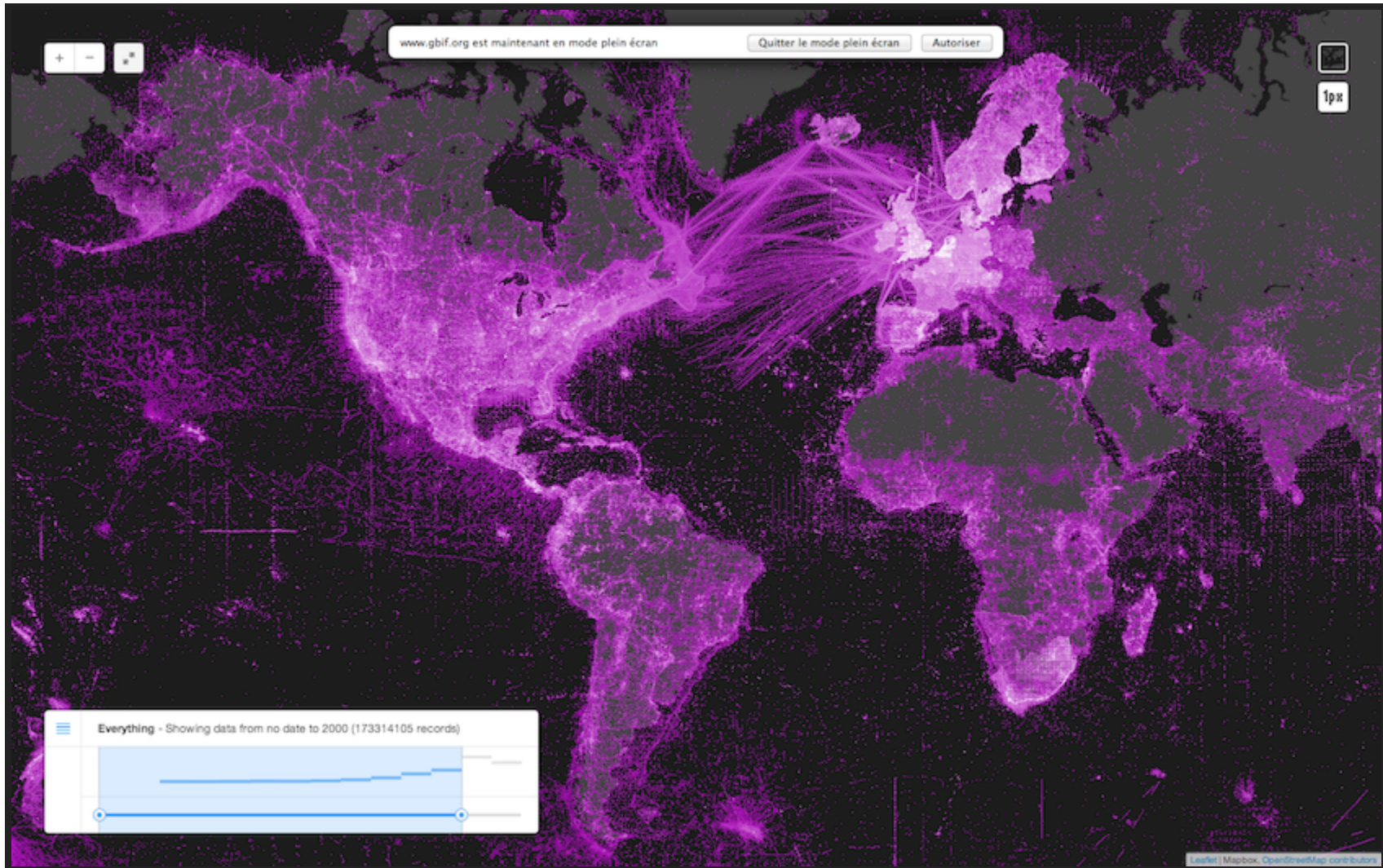


## 2. POURQUOI PUBLIER SES DONNÉES?

---

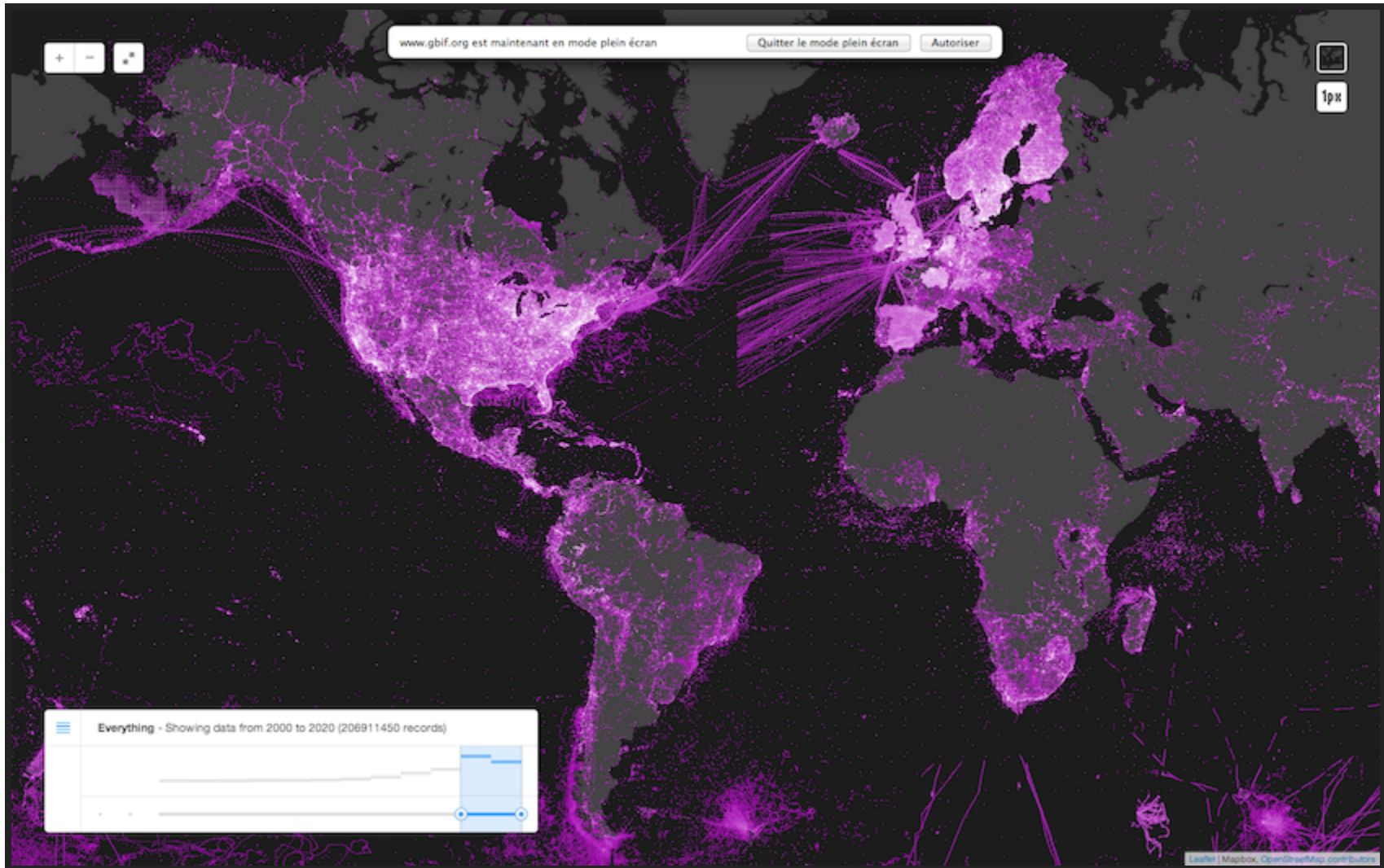
- Inventorier la Biodiversité du pays
- Comprendre son évolution et enrayer son déclin
- Décider sur base de connaissances scientifiques
- Renforcer la crédibilité des travaux/publications et augmenter la renommée des Auteurs et de leurs Institutions
- S'offrir de nouvelles opportunités de collaboration
- Permettre à de nouvelles recherches d'utiliser gratuitement les données publiques déjà recueillies
- Rapatrifier les connaissances dans leur pays d'origine
- Etre citer par la publication d'un **data paper**

## 2. POURQUOI PUBLIER SES DONNÉES?



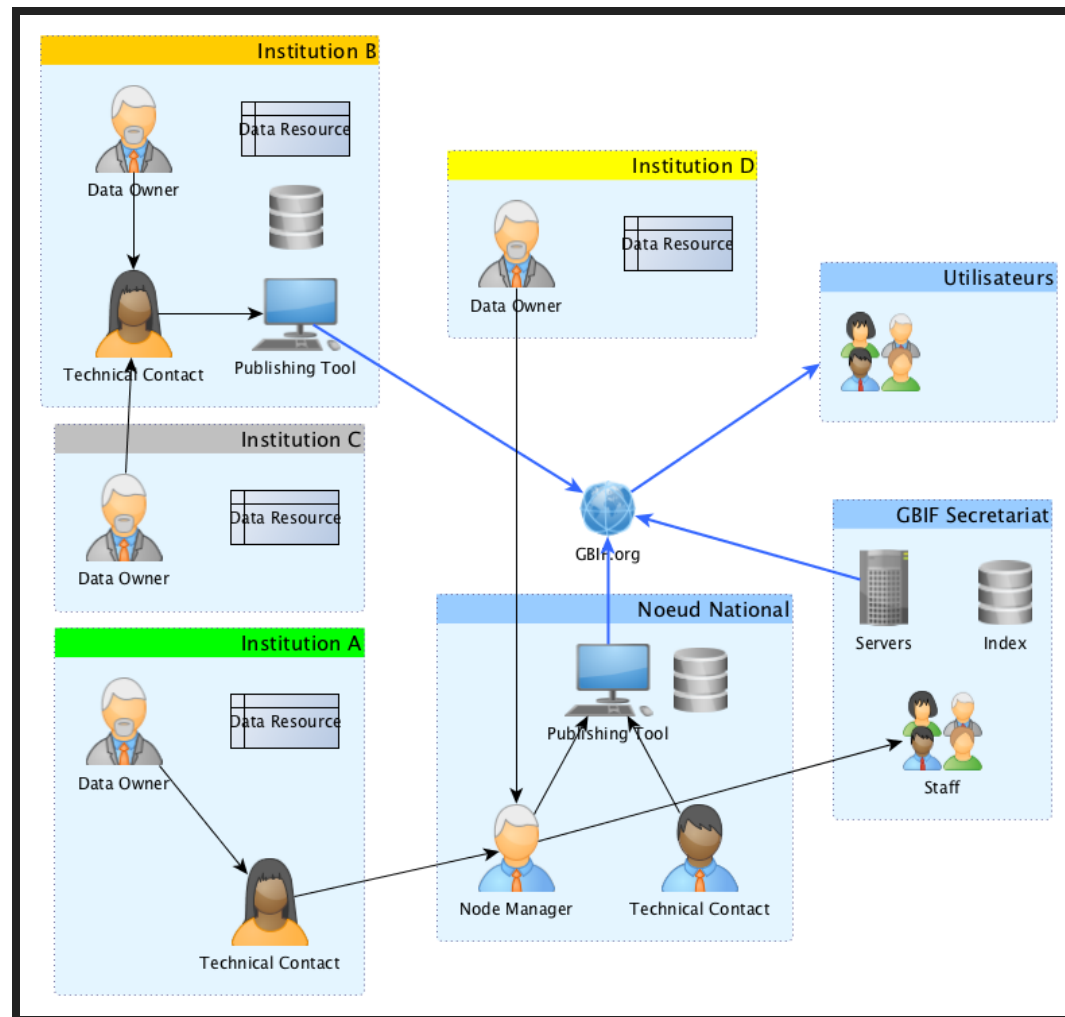
Données historiques XIX-XX  
173 millions d'occurrences, 25% de specimens

## 2. POURQUOI PUBLIER SES DONNÉES?



Données récentes XXI  
207 millions d'occurrences, < 4% de specimens

# 3. RÔLES DES INTERVENANTS



### 3.RÔLES DES INTERVENANTS

---

- Dans l'Institution:
  - Auteur(s) des données
  - (Contact Technique)
  - (Contact Administratif)
- Au Noeud national:
  - Gestionnaire du Noeud
  - Technicien(s)
- Au Secrétariat:
  - Développeurs
  - Support

# 4. PROPRIÉTÉ INTELLECTUELLE

---

Le GBIF encourage la libre diffusion des données sur la biodiversité et, plus particulièrement :

- ne revendique aucun droit de propriété sur le contenu des bases de données constituées par d'autres organismes et subséquemment affiliées au GBIF;
- cherche dans toute la mesure du possible à rendre librement et facilement accessibles, avec un minimum de restrictions à leur réutilisation, toutes les données commandées, créées ou constituées directement par le GBIF;
- respecte les conditions fixées par les éditeurs de données qui décident d'affilier leurs bases de données au GBIF

## 4. PROPRIÉTÉ INTELLECTUELLE

### MENTION DE LA SOURCE

---

- Le GBIF doit s'assurer que l'éditeur/détenteur des données est bel et bien mentionné et que les mentions de source sont maintenues lors des utilisations ultérieures des données.

## 4. PROPRIÉTÉ INTELLECTUELLE

### DROITS DE PROPRIÉTÉ INTELLECTUELLE SUR LES OUTILS APPLICABLES À LA BIODIVERSITÉ

---

- Le GBIF peut revendiquer les droits de propriété intellectuelle appropriés existants des instances nationales concernées sur tous les outils, comme des moteurs de recherche ou d'autres produits logiciels, conçus par le GBIF dans l'exécution de son programme de travail.



## 4. PROPRIÉTÉ INTELLECTUELLE

### TRANSFERT DE TECHNOLOGIES

---

- Les participants reconnaissent que, sous réserve des droits de propriété intellectuelle pertinents, le GBIF doit s'efforcer de promouvoir le transfert non exclusif, selon des modalités mutuellement convenues et à des établissements de recherche, surtout dans les pays en développement, des technologies informatiques qui sont disponibles, en particulier en ce qui concerne les programmes de formation et de renforcement des capacités.

# 5. QUE PEUT-ON PUBLIER?

---

- Des occurrences:
  - specimens de Collection
  - observations
- Des listes d'espèces (= checklists)
- Des meta-données seules  
(eg si les données sont pas numérisées)

voir [comment publier au GBIF?](#)

## 5. QUE PEUT-ON PUBLIER?

---

- des descriptions d'espèces
- des mesures
- des échantillons
- des documents multimédia
- des séquences ADN
- des liens vers d'autres sites

voir [les extensions au DarwinCore](#)

# 6. LES DONNÉES SENSIBLES

---

*“Qu'est-ce qu'une donnée sensible?”*

---

Toute donnée dont la publication est problématique:

- Pour l'espèce en question (eg braconnage)
- Pour l'observateur ou le collectionneur(eg législation)
- Pour l'auteur ou son institution  
(eg embargo sur résultats non publiés)

## 6.LES DONNÉES SENSIBLES

---

*“Que faire des données sensibles?”*

- Ne pas les publier (métadonnées seulement)
- Ne les publier que partiellement (les données non problématiques)
- Omission ou floutage des coordonnées(eg grille 10x10km)
- Identifier le genre plutôt que l'espèce



# 7. PRÉPARER LA PUBLICATION

```
région, localité, coordonnées, latitude, longitude, incertitude  
Aggui,Adrar,"21°10'N, 13°07'W",21.166666666666668,-13.116666666666667,1500  
Amogjar pass,Adrar,"20°32'N, 12°48'W",20.533333333333335,-12.8,1500  
Atar,Adrar,"20°31'N, 13°03'W",20.516666666666666,-13.05,1500  
Ain Savra,Adrar,"19°32.416'N, 12°06.829'W",19.540266666666668,-12.113816666666666  
Azougui,Adrar,"20°30'N, 13°10'W",20.5,-13.166666666666666,1500
```

```
{"usageKey":2440447,"scientificName":"Tursiops truncatus (Montagu, 1821)",  
"canonicalName":"Tursiops truncatus","rank":"SPECIES","synonym":false,  
"confidence":100,"matchType":"EXACT","kingdom":"Animalia","phylum":"Chordata",  
"order":"Cetacea","family":"Delphinidae","genus":"Tursiops","species":"Tursiops  
"kingdomKey":1,"phylumKey":44,"classKey":359,"orderKey":733,"familyKey":5314,  
"genusKey":2440446,"speciesKey":2440447,"class":"Mammalia"}
```

```
<!--?xml version="1.0" encoding="utf-8" ?-->  
<kml xmlns="http://www.opengis.net/kml/2.2">  
<document><folder><name>LRMgazetteer</name>  
<schema name="LRMgazetteer" id="LRMgazetteer">  
<simplefield name="Name" type="string"></simplefield>  
<simplefield name="Description" type="string"></simplefield>  
...  
</schema>  
<placemark>  
<name>Amogjar pass</name>  
<description>Adrar</description>
```

...

```
</placemark>  
</folder></document></kml>
```



## 7. PRÉPARER LA PUBLICATION

# PRÉ-REQUIS

---

Les données à publier doivent être:

1. sous forme électronique et dans un format exploitable
2. structurées et compréhensibles par une machine  
(eg CSV, JSON, SQL, XML)
3. expliquées & documentées
4. nettoyées & validées
5. pré-filtrées (ou filtrables) pour la publication
6. identifiables de manière unique et durable

---

*“C'est la partie la plus laborieuse du processus de publication!”*

## 7. PRÉPARER LA PUBLICATION

# OCCURRENCE\_IDS

---

Les bonnes pratiques à suivre:

1. Ils doivent être uniques dans le jeu de données
2. Ils doivent être stables dans le temps
3. Ils devraient être uniques globalement  
(eg 'urn:MRBIF:ENS:HNM:040661')

voir [IPT v2.1- Promoting the use of stable occurrenceIDs](#)

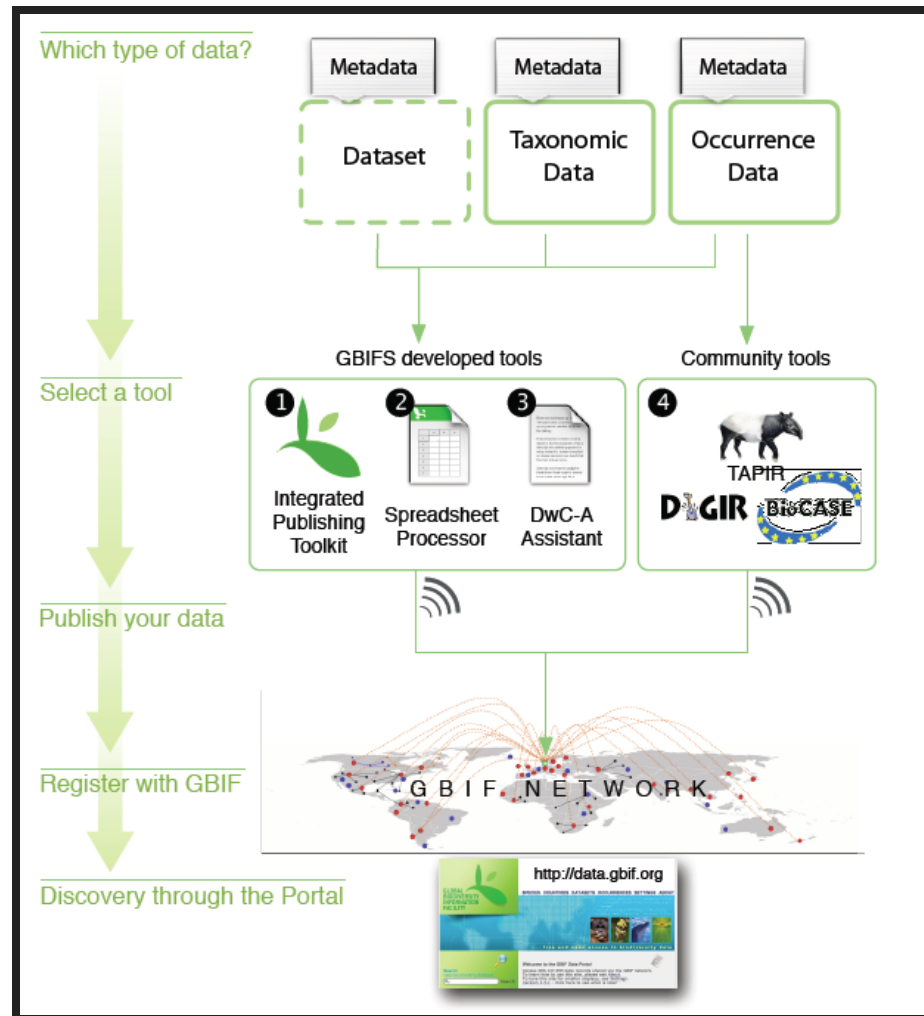
## 7. PRÉPARER LA PUBLICATION

# LES OUTILS RECOMMANDÉS

---

- un bon éditeur de texte (eg [Atom](#), [Emacs](#), [jEdit](#), [Vim](#))
- [QuantumGIS](#)
- [SQLite](#)
- [LibreOffice](#)
- [OpenRefine](#)
- les [webservices](#) du GBIF

# 8.COMMENT PUBLIER?



## 8.COMMENT PUBLIER?

# OUTILS RECOMMANDÉS

---

- Outils développés par GBIF:
  - [IPT](#) (Integrated Publishing Toolkit)
  - DarwinCore Assistant (2011)
  - Spreadsheet Processor (2011)
- Les autres outils:
  - [BioCAsE](#)
  - Tapir (2010)
  - Digir (2004)

## 8.COMMENT PUBLIER?

# IPT

---

*“Démonstration de publication des données avec l'Integrated Publication Toolkit”*

---

voir [IPT v2.1- Manuel d'utilisateur](#)

## 8.COMMENT PUBLIER?

# LES MÉTADONNÉES

---

- **Base**: Titre, Description, Type, Langue, Contact
- **Géo**: Coordonnées, Description
- **Taxo**: Nom scientifique, Rang, Description
- **Tempo**: Dates, Type
- **Mots-Clés**
- **Projet**
- ...

## 8.COMMENT PUBLIER?

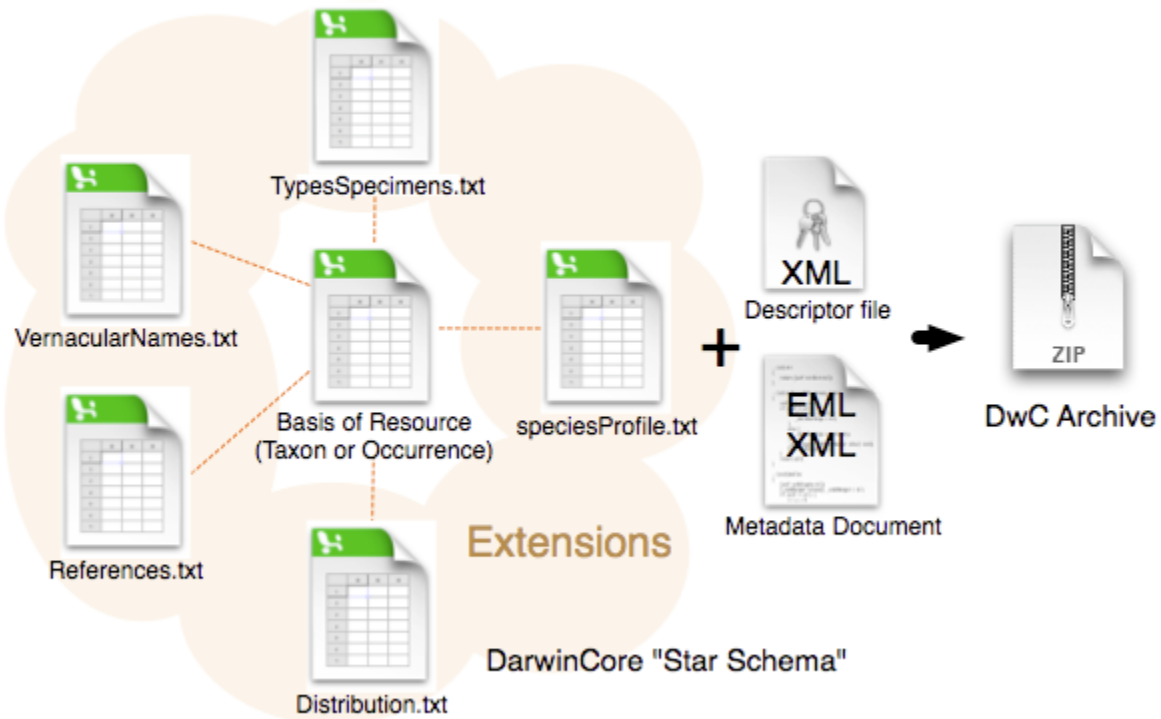
# LES DONNÉES

---

- Type: Occurrences, liste d'espèces ou metadonnées seules
- Conversion (quasi-automatique) en [DarwinCore](#)



## Darwin Core Archive Components



# 9. ACCÈS AUX DONNÉES

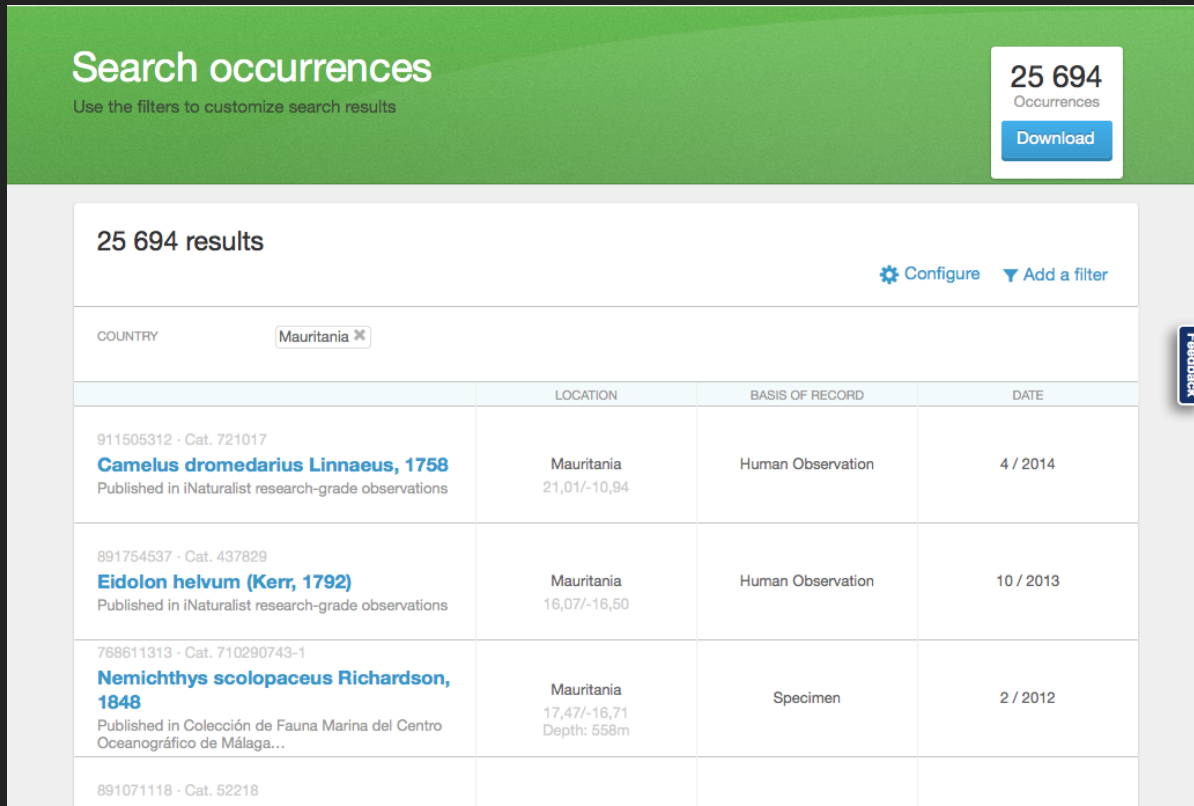
The screenshot shows the GBIF homepage with a green background and a world map. The main title is 'Global Biodiversity Information Facility' with the subtitle 'Free and Open Access to Biodiversity Data'. Navigation links include 'Data', 'News', 'Community', and 'About'. A 'Login or Create a new account' link is in the top right. A 'Feedback' button is on the right side. The statistics section displays: 449,696,555 OCCURRENCES, 1,454,694 SPECIES, 15,321 DATASETS, and 626 DATA PUBLISHERS. Three main content areas are: 'Sharing biodiversity data for re-use' (with links: Learn about GBIF, Publish your data through GBIF, Technical infrastructure), 'Providing evidence for research and decisions' (with links: Using data through GBIF, Enabling biodiversity science, Supporting global targets), and 'Collaborating as a global community' (with links: Current Participants, How GBIF is funded, Enhancing capacity). A search bar at the bottom contains the text 'Search news items and information pages...' and a 'Search' button.

Pré-requis:

1. Créer un compte sur [www.gbif.org](http://www.gbif.org)
2. Accepter l'accord d'utilisation

# 9.ACCÈS AUX DONNÉES

## RECHERCHER DES DONNÉES



The screenshot displays the 'Search occurrences' interface on a green background. It shows a search for 'Mauritania' with 25,694 results. A 'Download' button is visible. Below the search bar, there are options to 'Configure' and 'Add a filter'. The results are presented in a table with columns for 'COUNTRY', 'LOCATION', 'BASIS OF RECORD', and 'DATE'. The first three rows show detailed information for specific species: *Camelus dromedarius*, *Eidolon helvum*, and *Nemichthys scolopaceus*.

**Search occurrences**  
Use the filters to customize search results

25 694 Occurrences  
[Download](#)

25 694 results [Configure](#) [Add a filter](#)

COUNTRY

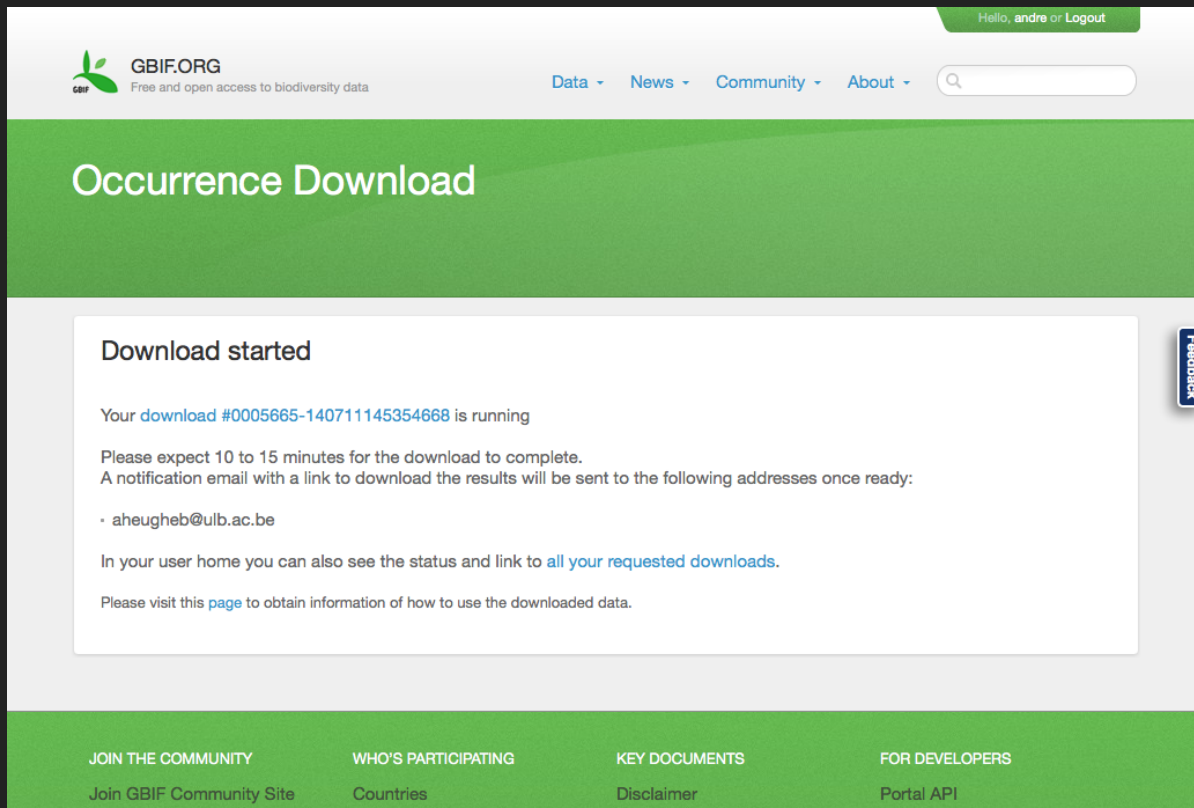
	LOCATION	BASIS OF RECORD	DATE
911505312 · Cat. 721017 <b>Camelus dromedarius Linnaeus, 1758</b> Published in iNaturalist research-grade observations	Mauritania 21,01/-10,94	Human Observation	4 / 2014
891754537 · Cat. 437829 <b>Eidolon helvum (Kerr, 1792)</b> Published in iNaturalist research-grade observations	Mauritania 16,07/-16,50	Human Observation	10 / 2013
768611313 · Cat. 710290743-1 <b>Nemichthys scolopaceus Richardson, 1848</b> Published in Colección de Fauna Marina del Centro Oceanográfico de Málaga...	Mauritania 17,47/-16,71 Depth: 558m	Specimen	2 / 2012
891071118 · Cat. 52218			

[Feedback](#)

Exemple: 25.694 occurrences pour la Mauritanie

# 9. ACCÈS AUX DONNÉES

# TÉLÉCHARGER DES DONNÉES



The screenshot shows the GBIF website interface. At the top left is the GBIF logo with the text "GBIF.ORG" and "Free and open access to biodiversity data". To the right are navigation links for "Data", "News", "Community", and "About", along with a search bar. A user greeting "Hello, andre or Logout" is visible in the top right corner. The main heading is "Occurrence Download". A central white box contains the following text:

**Download started**

Your [download #0005665-140711145354668](#) is running

Please expect 10 to 15 minutes for the download to complete.  
A notification email with a link to download the results will be sent to the following addresses once ready:

- [aheugheb@ulb.ac.be](mailto:aheugheb@ulb.ac.be)

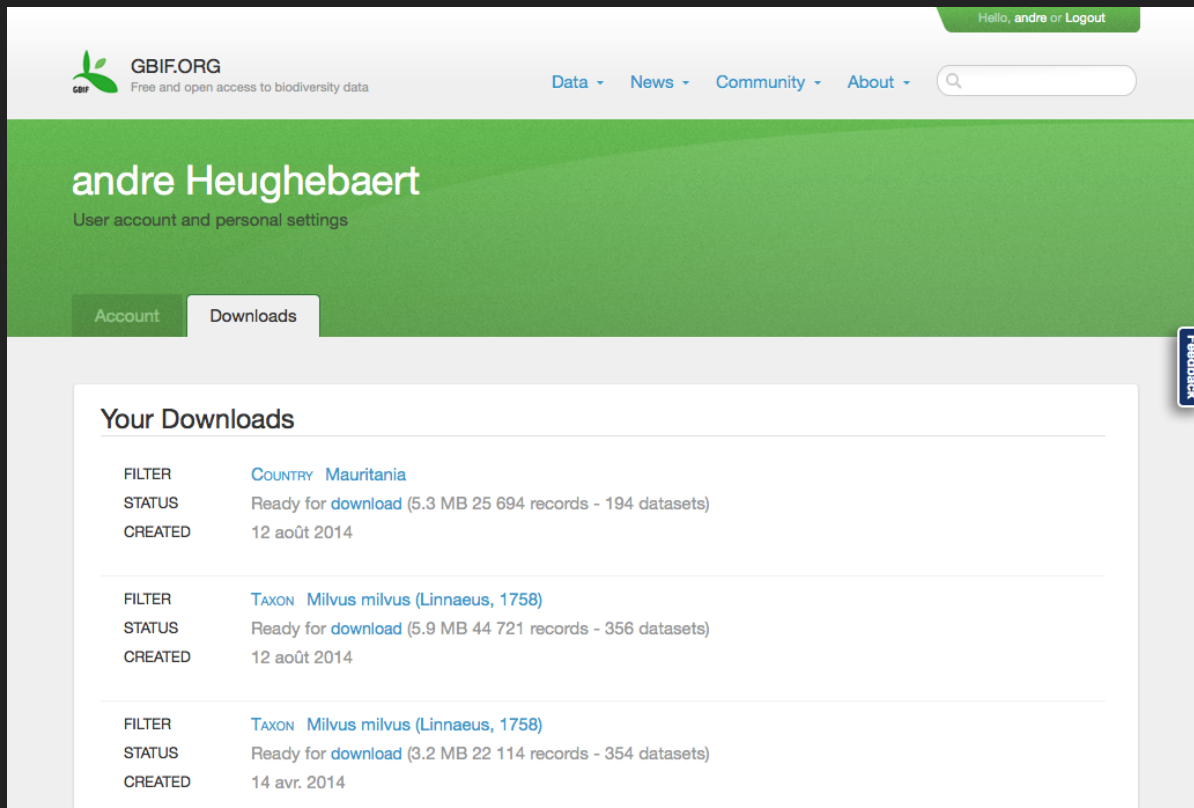
In your user home you can also see the status and link to [all your requested downloads](#).

Please visit this [page](#) to obtain information of how to use the downloaded data.

A "Feedback" button is located on the right side of the white box. The footer contains four columns of links: "JOIN THE COMMUNITY" (with sub-link "Join GBIF Community Site"), "WHO'S PARTICIPATING" (with sub-link "Countries"), "KEY DOCUMENTS" (with sub-link "Disclaimer"), and "FOR DEVELOPERS" (with sub-link "Portal API").

# 9. ACCÈS AUX DONNÉES

## REVOIR SES TÉLÉCHARGEMENTS



The screenshot shows the GBIF user account page for 'andre Heughebaert'. The page is divided into 'Account' and 'Downloads' tabs. The 'Downloads' tab is active, displaying a list of three download records. Each record includes a filter (Country or Taxon), status (Ready for download), size, number of records and datasets, and the creation date.

GBIF.ORG  
Free and open access to biodiversity data

Hello, andre or Logout

Data - News - Community - About -

andre Heughebaert  
User account and personal settings

Account Downloads

### Your Downloads

FILTER	COUNTRY <a href="#">Mauritania</a>
STATUS	Ready for <a href="#">download</a> (5.3 MB 25 694 records - 194 datasets)
CREATED	12 août 2014

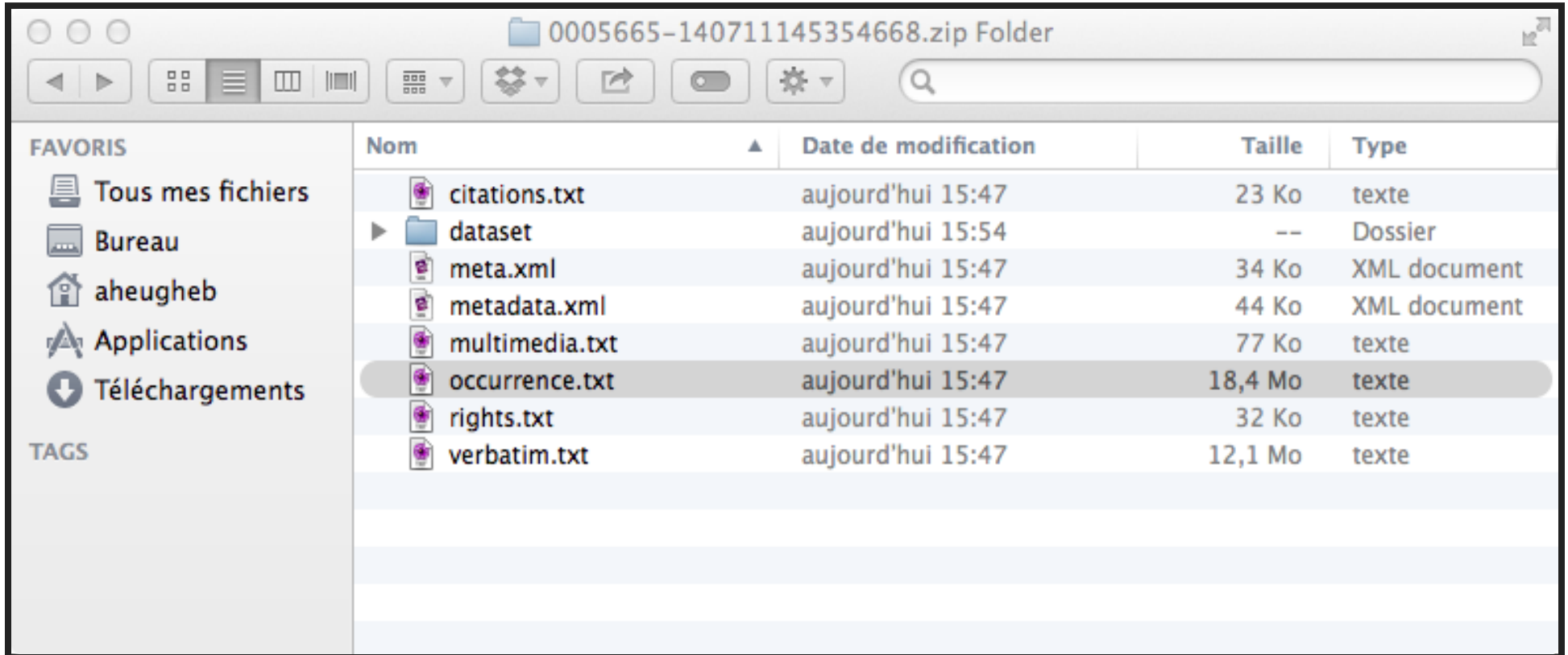
FILTER	TAXON <a href="#">Milvus milvus (Linnaeus, 1758)</a>
STATUS	Ready for <a href="#">download</a> (5.9 MB 44 721 records - 356 datasets)
CREATED	12 août 2014

FILTER	TAXON <a href="#">Milvus milvus (Linnaeus, 1758)</a>
STATUS	Ready for <a href="#">download</a> (3.2 MB 22 114 records - 354 datasets)
CREATED	14 avr. 2014

Feedback

# 9.ACCÈS AUX DONNÉES



Nom	Date de modification	Taille	Type
citations.txt	aujourd'hui 15:47	23 Ko	texte
dataset	aujourd'hui 15:54	--	Dossier
meta.xml	aujourd'hui 15:47	34 Ko	XML document
metadata.xml	aujourd'hui 15:47	44 Ko	XML document
multimedia.txt	aujourd'hui 15:47	77 Ko	texte
occurrence.txt	aujourd'hui 15:47	18,4 Mo	texte
rights.txt	aujourd'hui 15:47	32 Ko	texte
verbatim.txt	aujourd'hui 15:47	12,1 Mo	texte

- **Citations.txt**: Comment citer les sources des données?
- **Occurrence.txt**: Données indexées par GBIF
- **Rights.txt**: Eventuelles restrictions d'utilisation
- **Verbatim.txt**: Données telles que publiées par l'éditeur

# 10.UTILISATIONS

---

Recherche basée sur des données extraites du GBIF:

- Quelques exemples sur le [portail du GBIF](#)
- Les articles scientifiques via [Mendeley GBIF Public Library](#)  
(eg [publications d'auteurs belges](#))

# 10.UTILISATIONS

---

La publication des données offre de nouvelles perspectives.

- Au niveau **national**, elle décroïsonne les chercheurs et les disciplines, les départements et les institutions.
- Au niveau **international**, elle permet de mieux gérer certaines problématiques (eg espèces migratrices, espèces invasives et zones frontalières).



# MERCI POUR VOTRE PARTICIPATION

André Heughebaert, [Noeud Belge du GBIF](#)

[a.heughebaert@biodiversity.be](mailto:a.heughebaert@biodiversity.be)



Fait avec [reveal.js](#) - HTML Presentations made easy