



## RAPPORT DE MISSION



Point Nodal GBIF

Abdoulaye Biné GUINDO  
Ingénieur des Eaux et forêts

Décembre 2016

## SOMMAIRE

<b>I. Introduction</b> .....	<b>3</b>
<b>II. Déroulement de l'atelier</b> .....	<b>3</b>
2.1. Cérémonies d'ouverture et de clôture de l'atelier .....	3
2.2. Présentation des modules.....	3
2.2.1. Présentation des objectifs de l'atelier .....	3
2.2.2. Présentation du Système Mondial d'Accès à l'Information sur la Biodiversité GBIF .....	3
2.2.3. Les principes de gestion des données dans le processus de digitalisation .....	4
2.2.3.1. Principes de la publication.....	4
2.2.3.2. Flux des données: Vision des fournisseurs .....	4
2.2.3.3. Caractéristiques des données.....	6
2.2.3.3.1. Données Taxonomiques .....	6
2.2.3.3.2. Données spatiales.....	6
2.2.3.3.3. Données de collecte.....	7
2.2.3.3.4. Fournir la documentation sur les données .....	7
2.2.3.3.5. Flux des données: Chaîne des données .....	8
2.2.4. Utilité de la Biodiversité Informatique .....	8
2.2.5. Capture de données d'images, utilisation de BRAHMS .....	9
2.2.6. Nettoyage et formatage des données .....	10
2.2.7. Open Refine.....	12
2.2.8. Principes de géo référencement/utilisation de GEOlocate .....	12
2.2.9. Publication des données de la biodiversité .....	13
2.2.10. Patrimoine des bibliothèques de la biodiversité (BHL) .....	14
2.2.11. Digitalisation et présentation des données au format adéquat.....	14
2.2.12. Conversion des coordonnées géographiques en degré décimal.....	16
2.2.13. Utilisation d'Open Refine.....	16
2.2.13.1. Le facetage .....	17
2.2.13.2. Le filtrage .....	17
2.2.13.3. Le regroupement.....	17
<b>III. Conclusion</b> .....	<b>17</b>

## I. Introduction

Il a été constaté que la dégradation de la biodiversité connaît un seuil croissant au niveau de plusieurs pays africains. Pour sauver la biodiversité des mesures de conservation doivent être prises à partir des informations et des données fiables sur la biodiversité.

L'atelier régional de renforcement des capacités et de mobilisation des données de la biodiversité pour la conservation, l'utilisation durable et les prises de décisions en Afrique et à Madagascar qui s'est déroulé à Abomey au Bénin a porté sur la mobilisation des données de biodiversité, leur saisie, leur formatage, leur nettoyage et leur publication.

L'atelier a permis aussi aux participants de comprendre les missions du Système Mondial d'Accès à l'Information sur la Biodiversité (GBIF), les conditions d'adhésion au GBIF, les avantages d'appartenir au GBIF, etc.

## II. Déroulement de l'atelier

L'atelier régional de renforcement des capacités et de mobilisation des données de la biodiversité pour la conservation, l'utilisation durable et les prises de décisions en Afrique et à Madagascar s'est tenu du 28 Novembre au 2 Décembre 2016 à la Maison d'Accueil Sainte Anouarite d'Abomey Calavi, Bénin. Il a été marqué par :

### 2.1. Cérémonies d'ouverture et de clôture de l'atelier

Elles ont été caractérisées par les interventions du Doyen de la Faculté des Sciences Agronomiques de l'Université d'Abomey Calavi, du Chef de la Délégation Nationale GBIF Bénin et du Coordinateur Régional du Consortium GBIF Afrique.

Ils ont remercié les partenaires techniques et financiers et ont insisté sur le renforcement des capacités en ressources humaines, matérielles/équipements et financières des institutions pour la conservation et la valorisation de la biodiversité, socle du développement durable.

### 2.2. Présentation des modules

#### 2.2.1. Présentation des objectifs de l'atelier

L'objectif de l'atelier a consisté à former les participants sur les techniques avancées de collecte des données de la biodiversité, de contrôle de leur qualité, leur formatage, leur nettoyage et leur publication sur les sites du GBIF et celui de la section GBIF du Mali.

#### 2.2.2. Présentation du Système Mondial d'Accès à l'Information sur la Biodiversité GBIF

Le Système Mondial d'Accès à l'Information sur la Biodiversité ou Global Biodiversity Information Facility (GBIF) est un réseau mondial créé en 2001 sur initiative de l'Organisation de Coopération et de Développement Économiques (OCDE) en vue de définir des stratégies et des actions de conservation de la biodiversité.

Le GBIF a pour mission de faciliter l'accès libre par internet aux données primaires de la biodiversité pour soutenir la recherche.

Le GBIF dispose actuellement dans sa base de données de près de 640 millions de données d'occurrence provenant de plus de 1,6 million d'espèces végétales, animales, fongiques...

Il compte 38 pays votants, 19 pays associés et 39 organisations internationales.

L'Afrique dont le Mali fait partie dispose de moins de 4% de ces données et connaît aussi un manque crucial de compétences scientifiques et techniques en biodiversité. Le nombre de pays africains membres du GBIF est très faible (13).

Des dispositions doivent être prises pour que le Mali puisse adhérer au GBIF en qualité de pays associé pour une période de 5 ans sans contribution financière.

Les participants sont venus des huit pays africains membres du Consortium (Benin, Côte d'Ivoire, Guinée Conakry, Mali, Madagascar, Niger, Sénégal et RD Congo).

L'accent a été mis sur la vision du GBIF : encourager la publication de données les plus riches possibles pour assurer leur utilisation dans le plus large éventail d'approches et questions scientifiques.

### **2.2.3. Les principes de gestion des données dans le processus de digitalisation**

La communication sur les principes de gestion des données dans le processus de digitalisation a traité les notions relatives aux flux des données, la vision du GBIF, la chaîne des données, la qualité des données, les tâches de l'équipe de numérisation et la fourniture de la documentation sur les données.

#### **2.2.3.1. Principes de la publication**

Les principes de la publication des données de la biodiversité passent par les activités suivantes :

- Assembler des données de diverses sources ;
- Transformer des bases de données existantes ;
- Faire des traitements additionnels pour adapter les informations à publier aux standards de publication ;
- Contrôler la qualité des données obtenues auprès d'un fournisseur ;
- Prévenir les erreurs que d'essayer de les trouver et de les corriger (Chapman, 2005).

***Il faut cependant souligner que l'identification des erreurs et le nettoyage très important***

#### **2.2.3.2. Flux des données:** Vision des fournisseurs

Le flux des données de la biodiversité repose sur des normes qui sont entre autres :

- Utiliser les standards (exemple: le spreadsheet) ;
- Réfléchir à long terme ;
- Tenir compte des besoins des utilisateurs ;
- Investir dans la documentation et les métadonnées ;
- Contrôler la Qualité des données: Nettoyage des données car toute base de données contient des erreurs, on ne peut y échapper ! (Chapman, 2005).



Exemple spécimen image d'herbier

Il faut toujours prévoir un Champ d'erreur de 1 à 5 % à prévoir (Redman, 1996). Ce qui compte c'est d'avoir connaissance de ces erreurs et de savoir si elles restent dans des limites acceptables au regard de ce que l'on veut en faire.

***Pour ce faire l'équipe de numérisation doit s'assurer de la meilleure qualité possible lors de la numérisation des données taxonomiques, des données géographiques, des données sur la collection et le collecteur, des données descriptives et aider à documenter le(s) jeu(x) de données grâce aux métadonnées et annotations sur les occurrences.***



Saisie des données contenues dans les dossiers

### 2.2.3.3. Caractéristiques des données

#### 2.2.3.3.1. Données Taxonomiques

Elles portent sur :

- les noms (scientifique, vernaculaire, rang, hiérarchie, ...) ;
- les statuts (synonymes, nom valide, ...);
- les références (auteur, date et lieu) ;
- la détermination (par qui et quand ?) ;
- les champs relatifs à la qualité (certitude, ...).

Les erreurs sur les données taxonomiques portent généralement sur :

- les noms scientifiques et les noms communs à tous les niveaux de taxonomie.
- identifications incorrectes (requiert l'aide d'un taxonomiste) ;
- les erreurs orthographiques (nettoyage des données) ;
- le mauvais format (nettoyage des données) ;
- l'erreur liée à la classification (Mise à jour des données, GBIF).

***Le nom, point d'entrée, mal renseigné, est un risque de propagation des erreurs tout au long du processus de publication des données.***

#### 2.2.3.3.2. Données spatiales

Les données spatiales (textuelles ou géo référencées) représentent un des aspects cruciaux pour déterminer l'adéquation à l'usage des données primaires de biodiversité à travers:

- la modélisation de la distribution des espèces ;
- les sélections des zones à protéger ;
- la gestion de l'environnement et des ressources ;



**Utilisation carte pour les coordonnées géographiques**

- la latitude et longitude ;
- l'aire ;
- le point + rayon
- la boîte englobante (bounding box = rectangle calculé à partir des coordonnées de deux points)
- le polyline ;
- la référence de grille ;
- le datum: système géodésique ;



### **GPS et cartes nécessaires pour la mobilisation des données**

Les erreurs sur les données spatiales portent entre autres sur l'inversion des coordonnées et SRS inadapté.

Les données spatiales à éviter sont les valeurs zéro, le datum inconnu et les problèmes de conversion.

#### **2.2.3.3.3. Données de collecte**

La collecte des données de la biodiversité est une étape importante dans le processus de mobilisation des données, leur saisie et leur publication. La collecte portent sur :

- le nom du collecteur ;
- la date de collecte ;
- les informations supplémentaires: habitat, sol, conditions météorologiques.

La pertinence dépend du type de jeu de données:

- Collection statique (musée) : nom et ID du collecteur, date, habitat, méthode de capture ...
- Échantillonnage et inventaires exhaustifs : +méthode, taille de la grille, fréquence, si des spécimens de référence ont été collecté (+références)

#### **2.2.3.3.4. Fournir la documentation sur les données**

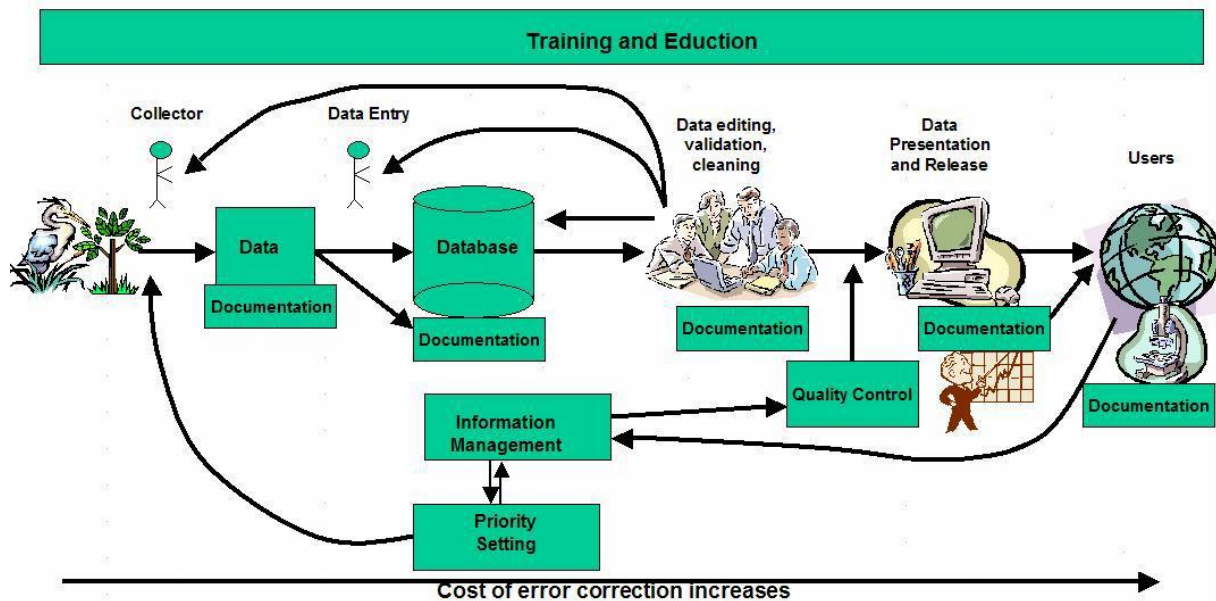
La fourniture de la documentation sur les données concerne :

- les métadonnées = «Données sur les données» ;
- le dataset ;
- la description du contenu, l'accessibilité, la complétude, ...
- la documentation de l'erreur ;
- la documentation des procédures de validation, de nettoyage et de correction appliquées.

***Les métadonnées doivent être suffisamment riches pour permettre l'usage des données par des tiers sans devoir se référer à la source de ces données.***

**Chaque donnée est associée à un unique DOI: citation des ressources par les utilisateurs**

### 2.2.3.3.5. Flux des données: Chaîne des données



Le flux des données ou chaîne des données comprend l'ensemble du processus de la collecte des données à la documentation (mise à disposition des utilisateurs des données de la biodiversité).

**Pendant le processus se rappeler que la perte de qualité des données peut survenir à chaque étape, que la responsabilité en terme de qualité de données doit être assignée le plus tôt possible dans cette chaîne.**

*Il faut noter aussi que la notion « qualité des données » est un concept relatif qui dépend de l'usage qui est fait de ces données et ne peut être jugé que par l'utilisateur.*

Dans une base de données, les données n'ont pas de qualité ou de valeur intrinsèque (Dalcin2004) ; elles n'ont qu'une valeur potentielle qui se réalise seulement quand quelqu'un utilise des données pour faire quelque chose d'utile.

**La qualité de l'information est liée à sa capacité à satisfaire ses consommateurs et leurs besoins (English 1999).**

### 2.2.4. Utilité de la Biodiversité Informatique

La biodiversité informatique ou l'informatique appliquée à la biodiversité est une nouvelle discipline en Afrique. Elle offre les opportunités de mobiliser des informations et des données sur la biodiversité, leur saisie, leur digitalisation, leur nettoyage, leur formatage, leur publication à travers des informations sur :

- Connaissance de la distribution des espèces à l'échelle des pays ;
- Connaissance de la distribution des espèces :
- Connaissance des facteurs du milieu qui régissent la distribution de l'espèce ;



- Facteurs environnementaux climatiques (pluviosité, humidité relative de l'air, la température...);
- Facteurs environnementaux non climatiques (sol, topographie, pression démographique, dégradation des habitats...);
- Facteurs non environnementaux (compétition entre espèces, barrières à la dispersion des espèces, dynamique des populations...);
- Connaissance de la niche écologique des espèces ;
- La niche écologique d'une espèce représente les conditions environnementales et biotiques dans lesquelles la population de l'espèce prospère indéfiniment sans contrainte de migration (Hutchinson, 1957) :
  - Prévisions de l'évolution des niches dans le contexte des changements globaux ;
  - Les facteurs environnementaux et non environnementaux qui caractérisent les niches des espèces évoluent et peuvent être projetés dans un futur plus ou moins lointain dans le contexte des changements climatique et globaux ;
  - Ceci permet d'identifier et de définir les zones favorables et non favorables à l'existence de l'espèce dans le présent et dans un futur plus ou moins lointain et ainsi, arrêter des stratégies de conservation des espèces à l'échelle.

A travers ses activités, le GBIF fait la promotion de la BI au service du développement durable.

La BI est un champ relativement nouveau d'investigation en science qui s'occupe de la collecte, du traitement et de l'analyse des données d'occurrence de la biodiversité et de la modélisation de la distribution des espèces (Peterson et al. 2010) pour la conservation et l'utilisation durable des services de la biodiversité.

La biodiversité informatique contribue à identifier et définir les stratégies adéquates de conservation des ressources naturelles et d'utilisation durable des services de la biodiversité dans le contexte des changements climatiques.

**Le soutien des décideurs politiques, des chefs d'institutions, des étudiants, des chercheurs, des enseignants et du public en général est indispensable pour relever les défis de la publication des données et de formation en biodiversité informatique.**

### **2.2.5. Capture de données d'images, utilisation de BRAHMS**

Recherche Botanique et Système de Gestion d'herbier (Botanical Research and Herbarium Management System BRAHMS) est un logiciel de mobilisation des données de la biodiversité à partir :

- Spécimens d'herbier ;
- Travaux de terrain ;
- Jardins botaniques ;
- Banques de semences ;
- Littérature ;
- Images de spécimens d'herbier ;
- Etc.

BRAHMS permet aussi de :

- Cataloguer et de faire le curetage des collections ;
- Faire des recherches taxonomiques ;
- Préparer checklists et flores ;

- Analyser et représenter sur une carte la diversité végétale ;
- Publier les données en ligne ;
- Ecrire des commandes par le biais de l'application FoxPro ;
- Utiliser un code à barres pour nommer les fichiers.

Pour son installation et sa configuration :

- télécharger BRAHMS et dézipper (décompresser) de préférence sur le C : <http://herbaria.plants.ox.ac.uk/bol/>;
- télécharger une licence « simple utilisateur » : fichier brahmskey.MEM ;
- licence multi-utilisateurs, faire une correspondance aux fournisseurs.

### **2.2.6. Nettoyage et formatage des données**

La communication a porté sur les points ci-dessous :

- Définition de concepts
- Besoin de nettoyage des données
- Où sont les erreurs
- Prévenir les erreurs
- Cas pratiques.

#### **a) Nettoyage des données**

Le nettoyage est procédé utilisé pour déterminer les données inexactes, incomplètes, ou déraisonnables et pour améliorer leur qualité grâce à la correction des erreurs et des omissions détectées (Chapman, 2005). Le procédé peut comprendre des contrôles de format, des contrôles d'exhaustivité, des contrôles de logique, des contrôles de limites, l'identification des valeurs aberrantes (géographiques, statistiques, temporelles ou environnementales) ou d'autres erreurs, et l'évaluation des données par des experts du domaine (par exemple des spécialistes en taxonomie). La nécessité de nettoyage des données est centrée sur l'amélioration de la qualité des données pour les rendre "aptes à l'utilisation" par les utilisateurs grâce à la réduction des erreurs dans les données et l'amélioration de leur documentation et présentation. L'aptitude à l'utilisation des données tourne autour de la précision des données, leur exactitude et authenticité pour des utilisations spécifiques (Faith et al. 2013).

### **b) Où sont les erreurs?**

Les données primaires sur les espèces englobent toute une gamme de données:

- données des musées et d'herbiers ;
- données d'observation ;
- données multimédias ;
- données d'échantillonnage (basées sur des protocoles de recherche).

De nombreux relevés ne portent qu'une description générale de l'endroit où ils ont été recueillis et n'ont pas de coordonnées géographiques.

Les données d'observation, de multimédia et d'inventaires sont précieuses pour de nombreuses études, mais parce que le matériel de référence (spécimen d'herbier) est rarement présent, l'information taxonomique ou nomenclature est généralement moins fiable que pour les collections des musées documentés.

### **c) Prévenir les erreurs**

La prévention des erreurs est préférable à leur correction ultérieure (Chapman, 2005).

Des outils sont mis au point pour aider les institutions à ajouter des informations géographiques (GEOLocate, BioGeomancer, Georeferencing Calculator...) ou à détecter et corriger les erreurs (OpenRefine ...).

Pour prévenir et corriger les erreurs taxonomiques, des listes de noms de référence d'espèces (checklists) au plan mondial et régional sont en cours d'élaboration et permettent de réduire les erreurs (catalog of life, <http://www.catalogueoflife.org/>).

### **d) Les erreurs spatiales**

Il est important d'assurer une détermination précise de l'espèce observée ou enregistrée. Il est recommandé de se référer aux taxonomistes et documents taxonomiques adéquats à cette fin. Si les coordonnées géographiques sont saisies en même temps que les observations où les occurrences, il est important de faire attention lors de l'encodage des latitudes et longitudes de manière à éviter les erreurs d'encodage.

Si les coordonnées géographiques ne sont pas enregistrées en même temps que les données d'observation ou données d'occurrence, il est alors important d'enregistrer les localités où les données sont collectées avec des détails complets (pays, département, commune, arrondissement, village ...) de telle sorte que l'utilisation des outils de géo référencement puisse permettre de fournir des coordonnées adéquates aux données.

### **e) Erreurs nomenclaturales et taxonomiques**

Les noms sont la clé majeure pour accéder à l'information dans les bases de données primaires des espèces. Si le nom est erroné, l'accès à l'information par les utilisateurs sera difficile, voire impossible. La plus facile à corriger de ces erreurs porte sur les données nomenclaturales (les fautes d'orthographe).

Des listes de noms (et synonymes) sont les outils clés pour aider à cette tâche. De nombreuses listes existent déjà pour les régions et / ou des groupes taxonomiques, et celles-ci sont progressivement intégrées dans les listes mondiales :

- TNRs: <http://tnrs.iplantcollaborative.org/index.html>;

- Catalog of life: <http://www.catalogueoflife.org/listmatching/>;
- Il faut signaler que dans le cadre du nettoyage des données, référence est faite au logiciel Open Refine.

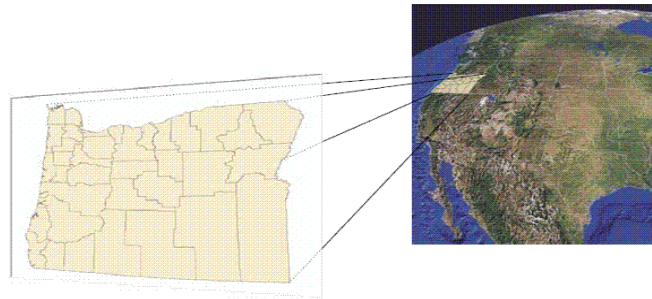
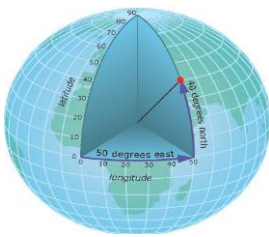
### 2.2.7. Open Refine

Open Refine est un outil qui permet aussi la saisie des données de la biodiversité à travers :

- Le chargement des fichiers et des projets de mobilisation des données qui peut se faire à partir de diverses sources de données comme TSV, CSV, SV, Excel (.xls et.xlsx), et des données XML de Google Docs. Le chargement des données comporte deux étapes : la première charger le fichier et la deuxième est de créer le projet ;
- Le facettage  
Il est une fonctionnalité qui permet d'obtenir un aperçu général des données et de filtrer les enregistrements que la personne veut voir ou à changer. Il facilite l'utilisation et l'analyse des données et peut être fait avec des cellules contenant tout type de texte, des chiffres et des dates.

### 2.2.8. Principes de géo référencement/utilisation de GEOlocate

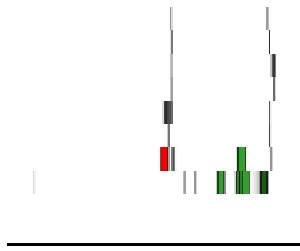
GEOlocate permet à partir des coordonnées cartographiques, affecter un emplacement spatial à des entités cartographiques.



Système de coordonnées géographiques: Latitude-Longitude

Les données spatiales utilisées dans le cadre de GEOlocate peuvent contenir des incertitudes liées :

- aux coordonnées ;
- à l'échelle de la carte ;
- à l'étendue de la localité ;
- à l'exactitude des données enregistrées par le GPS ;
- au datum ;
- à l'imprécision dans la mesure de l'Azimuth ;
- à l'imprécision dans la mesure des distances.



### Sources d'incertitudes : des points d'inventaire hors du territoire national (cas du Benin)

GEOLocate est une solution de géo référencement électronique complet conçu pour ces données, facile d'utilisation, réduit le temps et le coût requis pour géo référencer les données, disponible en ligne et hors ligne.



GEOlocate permet :

- d'importer des données de localisation existantes ;
- de traiter automatiquement des lots d'informations ;
- de projeter les données géo référencées avec possibilité de correction ;
- de calculer la marge d'erreur.

Georéférencement et GEOlocate ont des difficultés :

- Incertitude ;
- Niveau de connaissances du territoire ;
- Disponibilités des informations sur les différentes localités.

### **2.2.9. Publication des données de la biodiversité**

La publication des données de la biodiversité se fait à travers le logiciel GBIF Kit intégré de Publication IPT (Integrated Publication Toolkit) à travers les étapes suivantes :

- La création d'un compte ;
- La gestion des ressources (saisie des informations/données sur les espèces inventoriées) ;
- L'utilisation des métadonnées géographiques (couverture géographique);
- La correspondance des champs ;
- La cartographie ;
- La gestion de la ressource (rendre visible à tous) ;
- Le management de la ressource (ajouter un gestionnaire) ;
- La publication de la ressource (publier le jeu de données);
- Enregistrement de la ressource (moissonnage par GBIF international).

Pour amples informations consulter les références ci-dessous :

IPT v2 User Manual

<http://code.google.com/p/gbif-providertoolkit/wiki/IPT2ManualNotes>

Occurrence Data Publishing Tutorial

<http://code.google.com/p/gbif-providertoolkit/wiki/TutorialOccurrenceResource>

### **2.2.10. Patrimoine des bibliothèques de la biodiversité (BHL)**

Le Patrimoine des bibliothèques de la biodiversité a été créé en 2006 par un consortium de 10 au départ (32 actuellement) bibliothèques appartenant à des universitaires, des jardins botaniques et collections d'histoires naturelles.

L'objectif de la bibliothèque sur la biodiversité consiste à la digitalisation de la littérature sur la taxonomie à travers un « environnement ouvert » de l'Encyclopédie de la vie (Encyclopedia of life).

Financé en 2007 par Mac Arthur, les données et informations de BHL sont à accès libre et gratuit. Le secrétariat de BHL est basé à Washington DC à la bibliothèque Smithsonian.

Les missions de BHL consistent à l'amélioration de la recherche méthodologique à travers la collaboration afin que la littérature sur la biodiversité soit librement accessible en tant que élément de la communauté de la biodiversité mondiale.

BHL est consulté par plus de 178 000 personnes par mois, il renferme plus de 189 000 volumes en ligne, plus de 50 millions de pages et 111 000 titres, 36% de la littérature sur la biodiversité du domaine public sont disponibles.

BHL est composé :

- **Institutions membres** au nombre de 16 qui sont entre autres : la bibliothèque du Museum Américain d'histoire naturelle, BHL Australie, BHL Mexico, la bibliothèque de l'université de botanique de Harvard, jardin botanique de Kew, le jardin botanique de Missouri aux Etats Unis d'Amérique, bibliothèque de Smithsonian Etats Unis d'Amérique bibliothèque nationale de Singapour ;
- **Partenaires clés** : BHL Africa ; BHL Australia ; BHL China ; BHL Egypt ; BHL Europe ; BHL México ; BHL SciELO (Brasil) et BHLSingapore.

### **2.2.11. Digitalisation et présentation des données au format adéquat**

La présentation des données de biodiversité obéit à des principes et critères établis par GBIF qui sont entre autres :

#### **A) Attributs essentiels des données primaires**

La digitalisation des données primaires passe par trois attributs essentiels :

- Nom scientifique complet, univoque de l'organisme vivant observé/collecté, quel que soit le règne auquel il appartient ;
- Description complète du lieu d'occurrence (observation, collecte) de l'individu en question; si possible, coordonnées géographiques

– Date complète de collecte (jour, mois, année).

### **B) Darwin Core Archive**

Darwin Core Archive (DwC-A) est un standard de présentation des données de la biodiversité informatique qui utilise les termes de Darwin Core pour produire un seul jeu de données autonome pour les données d'occurrence ou les listes de contrôle (Checklists).

Essentiellement, il s'agit d'un ensemble de fichiers texte (CSV) avec un descripteur simple pour informer sur la façon dont les fichiers sont organisés.

Le format est défini dans les lignes directrices de texte Darwin Core. ([https://en.wikipedia.org/wiki/Darwin\\_Core\\_Archive](https://en.wikipedia.org/wiki/Darwin_Core_Archive))

**Il est le format préféré pour la publication de données sur le réseau GBIF, le spreadsheet de publication de données du GBIF en est dérivé.**

### **C) Principales composantes du spreadsheet**

Le spreadsheet comprend les feuilles des Métadonnées et les feuilles d'occurrences.

Les liens ci-dessous donnent amples informations sur le spreadsheet:

- TNRS: <http://tnrs.iplantcollaborative.org/index.html>;
- Catalog of life: <http://www.catalogueoflife.org/listmatching>;
- Canadensys coordinates conversion: <http://data.canadensys.net/tools/coordinates>;
- GPS Visualizer: <http://www.gpsvisualizer.com/>.

### **D) Les écueils à éviter**

Les écueils à éviter pendant la digitalisation des données sont :

- Données dupliquées qui sont des données relatives à la même espèce, observée à la même date et au même endroit ;

- Comment s'en rendre compte? En comparant les données dupliquées et les données d'abondance (Exemple de données recueillies sur les placettes d'inventaire). Les données primaires de la biodiversité Les données basées sur les protocoles de sondage sur les plantes au cours des inventaires de la végétation (liste des espèces affectées de coordonnées géographiques des points d'inventaires), le nombre d'individus par espèce représente une donnée d'abondance de l'espèce.

La présentation des jeux de données obéit à des règles de format dérivées du Darwin Core Archive qu'il faut respecter. Pour ce faire il faut :

- Présenter au complet les métadonnées ;

- Faire attention pour remplir correctement chaque colonne et chaque ligne de la feuille d'occurrence ;

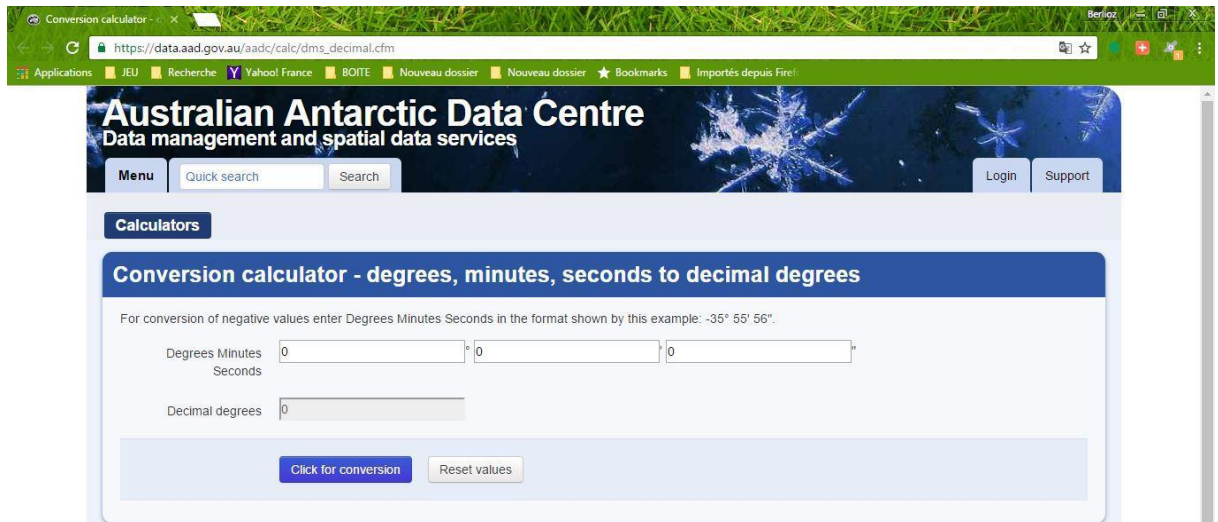
- Il ne faut pas hésiter à utiliser les outils disponibles sur les différents sites pour faire les conversions nécessaires et aussi pour nettoyer les données.

Toutes les étapes de saisie des données de biodiversité, leur formatage, leur nettoyage, le contrôle de leur qualité et leur publication se font grâce à la biodiversité informatique.

### 2.2.12. Conversion des coordonnées géographiques en degré décimal

Il s'agit de convertir les coordonnées géographiques (degrés, minutes et secondes) en degré décimal grâce à des méthodes de conversion disponibles:

- Les sites web



- Les formules mathématiques (Degré Minute Seconde -----> degré décimal) ;

$$Y = N 7^{\circ}23'24,0''$$

$$Y_d = 7 + (23/60) + (24/3600) = 7,39$$

$$X = E 2^{\circ}27'09,0''$$

$$X_d = 2 + (27/60) + (9/3600) = 2,4525$$

$$Y = S 7^{\circ}23'24,0'' = (-1 * (7 + (23/60) + (24/3600))) = -7,39$$

$$X = W 2^{\circ}27'09,0'' = (-1 * (2 + (27/60) + (9/3600))) = -2,4525$$

$$Y = S 7^{\circ}23'24,0'' = -7,39$$

$$X = E 2^{\circ}27'09,0'' = 2,4525$$

- Le Système d'Information Géographique SIG et QGIS ;
- Autres logiciels (dnr Garmin).

### 2.2.13. Utilisation d'Open Refine

Open Refine est un logiciel qui permet la saisie des données sur la biodiversité à travers des fonctionnalités ci-dessous. Il permet le chargement de fichiers et les projets à partir de diverses sources de données: TSV, CSV, SV, Excel (.xls et .xlsx), JSON, RDF et des données XML comme dans Google Docs. Le chargement des données comporte deux étapes: la première est de charger le fichier et la deuxième est de créer un projet de saisie des données de la biodiversité.



Il comprend les opérations suivantes :

#### **2.2.13.1. Le facettage**

Il est une fonctionnalité qui permet d'obtenir un aperçu général des données et de filtrer les enregistrements que nous voulons changer ou voir. Il facilite l'utilisation et l'analyse des données et peut être fait avec des cellules contenant tout type de texte, des chiffres et des dates ...

#### **2.2.13.2. Le filtrage**

Il est une fonctionnalité qui permet de faire des corrections sur les noms des genres, des espèces.

#### **2.2.13.3. Le regroupement**

Il permet d'avoir des informations et données sur le pays, les villes au niveau desquels la mobilisation des données sur la biodiversité a été faite.

### **III. Conclusion**

La mobilisation des données de la biodiversité est une opération indispensable pour toute politique de gestion durable des ressources naturelles.

Fort de cette logique, le Système Mondial d'accès à l'Information sur la Biodiversité accompagne le Mali dans la mise en œuvre du projet régional de renforcement des capacités, de mobilisation des données de la biodiversité pour leur conservation, leur utilisation durable et la prise de décisions en Afrique et à Madagascar.

Le processus de saisie des données de la biodiversité à partir des spécimens d'herbier, des travaux d'inventaire de terrain et des images qui démarré grâce à la **biodiversité informatique** doit être renforcé et appuyé.

L'adhésion du Mali au GBIF est une opportunité qui contribuera à la mise en œuvre à hauteur de souhait de la Politique Nationale de la Protection de l'Environnement et de la Stratégie Nationale de la Biodiversité.