



RAPPORT DE L'ATELIER DE RENFORCEMENT DE CAPACITES POUR L'ELABORATION ET LA PUBLICATION DES DATA PAPERS A PARTIR DES DONNEES DE LA BIODIVERSITE

Projet BID-AF2015-0004-NAC

INTRODUCTION

Dans le cadre de la mise en œuvre du projet « Biodiversity Information for Development » (BID), le nœud Togo du « Global Biodiversity Information Facility » (GBIF) placé sous la coordination de la Faculté des Sciences de l'Université de Lomé et géré avec l'appui administratif du Ministère de l'Environnement et des Ressources Forestières (MERF), a organisé du 12 au 14 décembre 2017 dans la salle de réunion de WASCAL de l'Université de Lomé, un atelier de renforcement de capacités en utilisation des données de la biodiversité en vue de produire des Data papers (articles à caractère scientifique) au profit des détenteurs et utilisateurs de données de la biodiversité au Togo.

L'objectif général de cet atelier est d'accroître la capacité du nœud et du réseau GBIF du Togo dans la publication de données en outillant les détenteurs et utilisateurs à l'élaboration et à la publication d'articles à caractères scientifiques à partir de leurs propres jeux de données.

L'objectif spécifique de cet atelier, est qu'à la fin de celui-ci, chaque participant ait soumis à un journal son propre "data paper" préparé et basé sur ses propres jeux de données, en utilisant IPT (Integrated Publishing Toolkit) du GBIF.

Les activités menées au cours de cet atelier sont :

- ✓ le traitement des données des participants ;
- ✓ la conversion d'un fichier Excel en fichier Texte et d'un fichier Texte en fichier TSV ;
- ✓ la génération des ressources au niveau de l'IPT (mapping) ;
- ✓ l'obtention de la version RTF ;
- ✓ l'amélioration du draft du data paper obtenu.

Ont pris part à cet atelier treize (13) personnes représentant les directions des ministères de l'environnement et des ressources forestières (direction des ressources forestières DRF), de l'agriculture, de l'élevage et de l'hydraulique (Institut togolais de recherche agronomique, ITRA), de l'urbanisme et du cadre de vie, de l'enseignement supérieur et de la recherche, les organisations non gouvernementales (ONG) en l'occurrence l'ONG Agbozégué. Les champs d'action de ces directions et ONG ont un lien avec la biodiversité. En outre, des enseignants-chercheurs et étudiants en biologie



UNION EUROPÉENNE



et informatique de l'Université de Lomé ont pris part à ces activités. La liste des participants est annexée au présent rapport.

Ce rapport est structuré en trois parties : l'introduction, le déroulement de l'atelier (cérémonie d'ouverture, le déroulement des travaux, la visite et impressions ; les recommandations ; et la Cérémonie de clôture) et la conclusion.

DEROULEMENT DE L'ATELIER

I - Cérémonie d'ouverture

L'atelier a été ouvert le mardi 12 décembre 2017 avec la cérémonie d'ouverture. Cette cérémonie a débuté à 9h 00 et a connu deux interventions, celles du responsable du nœud GBIF Togo, le Professeur RADJI et du formateur, Dr Francisco PANDO. Le responsable du nœud, a salué les participants pour avoir été sélectionné et les a remercié pour avoir répondu présents à l'invitation. Il a ensuite situé le contexte dans lequel s'inscrivent les objectifs dudit atelier en rappelant que le présent atelier est la suite d'une série d'activités organisés dans le cadre du projet BID au Togo. L'objectif principal de la rencontre de formation est d'outiller les détenteurs et utilisateurs des données de la biodiversité à l'élaboration et à la publication des articles scientifiques à partir de leurs propres données déjà publiées ou non sur le portail du GBIF. Il a par ailleurs insisté sur le fait que la présente session sera purement technique, pratique avec des exercices devant permettre aux participants d'utiliser les jeux de données qu'ils auront eux-mêmes retenues, pour faire l'objet d'une publication scientifique. Enfin, il a souhaité la cordiale bienvenue et remercié le formateur venu de l'Espagne à cet effet. Dr Francisco Pando pour sa part est revenu sur le caractère pratique de la formation et souhaité que les échanges soient ouverts entre participants et formateur pour la bonne maîtrise de la technique et des outils qui seront enseignés au cours de la formation. Il a par la suite soumis à l'appréciation des participants, le programme proposé pour l'atelier.

II - Déroulement des travaux

L'essentielle de l'atelier a consisté au déroulement de plusieurs modules, des discussions et des travaux pratiques.

1. Présentations

Toute la documentation pour cet atelier a été mise à la disposition des participants par le formateur à l'ouverture de la formation.

Lesdits documents restent accessibles sur Google Drive :

https://drive.google.com/drive/folders/1FwsV_FW3QSVK06MpJdXmMVzpDBDY9Ovk



UNION EUROPÉENNE



1.1. Introduction portant sur l'open science and open data, l'environnement électronique des données de biodiversité et le rôle du GBIF ainsi que l'importance de la publication des données au GBIF

Le formateur a fait une brève historique de l'évolution de la mobilisation des données ainsi que l'importance, le caractère des données pour les sciences. Il a parcourus un panel de sites avec leurs liens hébergeant les données portant sur la biodiversité notamment : Eol ; BHL ; IPNI ; TDGW ; GenBank ; Barcode of Life ; BOL & GenBank ; NCBI ; BOL ecosystem ; CETAF ; SciColl ; Life Watch ; IPBES ; etc.

1.2. Flux de publication de données à travers le GBIF

Les principales étapes du cycle des données sont entre autre : la collecte de données ; la préparation de base de données ; l'élaboration de métadonnées ; le téléchargement de données et de métadonnées ; la génération de data paper en fichier .rtf ; leur publication sur le www.gbif.org ; l'utilisation des données publiées ainsi que les révisions éventuelles.

1.3. Différentes licences de publication au GBIF

Les caractéristiques et l'utilisation des différentes licences (CC BY; CC BY-NC; CC BY-NC-SA; CC BY-NC-ND; CC BY-SA; CC BY-ND), les avertissements ; la citation pour les données disponibles ont été explicités.

1.4. Darwin Core standard (DwC)

C'est un standard utilisé pour la publication et le partage de données de biodiversité. Les références de ce standard ont été présentées. Ce standard créé en 1999 a évolué pour rendre la forme que l'on le connaît actuellement.

1.5. Darwin Core Archive (DwCA)

Le formateur a présenté les différences entre le DwCA et le DwC. En effet, DwC standard regroupe des données standards, une liste de termes et de définitions pour la publication et l'échange de données portant sur la biodiversité alors que le DwCA est simplement un outil de stockage électronique d'information pour héberger ou non des données en format Darwin Core standard.

1.6. Qualité, contrôle, conception et nettoyage des données

Cette présentation a porté sur les principes, la conception et les caractéristiques des données de qualité ; la définition et les fondements du nettoyage des données ainsi que de la qualité des informations géographiques et taxonomiques sur les données.



1.7. Data paper

Un data paper est défini comme un article spécifique décrivant des données particulières ou des collections de données. Il décrit donc des données et des faits avec des analyses. Un data paper doit répondre à un minimum d'exigences dont : la rigueur, la lisibilité, concision et précision, avoir une portée scientifique, etc. Les différentes parties d'un data paper ont été passées au crible à savoir : le titre, les auteurs, le résumé, les mots clés, les détails sur le projet, le contexte de collecte, la couverture taxonomique et géographique, la méthodologie, les tableaux, les graphiques, les résultats, la discussion, une conclusion et des références bibliographiques.

1.8. De l'IPT au data paper

L'IPT permet l'élaboration semi-automatique de data paper. En effet, l'IPT permet de renseigner et de compiler les différentes informations à renseigner dans les différentes parties d'un article. Cette application permet à terme l'obtention d'un draft structuré de data paper qui fera l'objet d'une amélioration et d'un enrichissement avec des cartes, graphiques et tableaux. Des exemples d'articles déjà parus ont été présentés ainsi que les journaux spécialisés (notamment Pensoft et autres) en Data Paper Publication avec leur facteur d'impact (IF).

2. Discussions

Les différentes présentations ont été suivies de débat pour une meilleure compréhension des participants.

3. Travaux pratiques

Les travaux pratiques ont permis à chaque participant de travailler sur ses propres données sous la supervision du formateur. Ces travaux pratiques ont concerné entre autre :

- ✓ l'harmonisation des titres des colonnes des données en Excel par inspiration du modèle proposé par le formateur ainsi que l'explication de ces termes ;
- ✓ le nettoyage des données ;
- ✓ la conversion des données du format Excel en format Texte ;
- ✓ la création de projet sous OpenRefine par importation de fichier Texte créé et son nettoyage ;
- ✓ l'enregistrement du fichier Texte assaini sous format .tsv ;
- ✓ l'ouverture de compte de chaque participant sur l'IPT Test ;
- ✓ l'importation du fichier .tsv dans l'IPT Test (au niveau de Manage ressources) ;
- ✓ le renseignement des champs correspondants sous DwC (mapping) ;
- ✓ le renseignement des différentes parties du data paper au niveau du champ metadata ;
- ✓ l'ouverture d'un aperçu du draft du data paper consolidé ;



- ✓ la visualisation sous format .rtf du draft du data paper ;
- ✓ les modifications possibles grâce à la fonction Edit.

III – Visite et impressions

Au troisième jour de l'atelier, les participants ont visité l'herbier de l'Université de Lomé qui fait office de l'Herbarium national.. Cette visite a permis aux participants de se familiariser avec les collections, de prendre connaissance de comment faire un herbier lors des travaux de terrain et de suivre la démonstration du processus de numérisation des spécimens.

Une évaluation orale a permis aux participants d'exprimer leurs impressions et sentiments sur l'atelier.

IV - Recommandations

A l'issue de l'atelier, il a été recommandé ce qui suit :

- les draft des data papers générés au cours de l'atelier doivent être complétés et améliorés en vue de leur publication. A cet effet, l'IPT Test sera maintenu actif jusqu'à la fin du mois de février 2018 pour permettre aux participants de continuer à s'entraîner ;
- le nœud GBIF Togo reste disponible pour apporter son appui et son assistance aux participants pour la finalisation de leur data paper. Le Node manager sera aidé par le comité de suivi du Réseau des détenteurs et utilisateurs des données de la biodiversité au Togo ;
- maintenir la dynamique et la solidarité au sein du groupe de participants à partir de la mailing liste créé ;

tous les participants seront co-auteurs des papiers à faire paraître si chacun fait l'effort de produire un draft sur ses propres données. La feuille de route proposée pour la suite des travaux (post-atelier) prévoit que les premiers draft incluant tous les aspects des metadonnées soient échangés avec le Node Manager et le formateur le 15 janvier 2018.

V - Cérémonie de clôture

L'atelier débuté le mardi 12 décembre à 9h a pris fin le jeudi 14 décembre 2017 à 17h 31 avec :

- la remise de l'attestation de participation à chaque participants;
- l'allocution du responsable de GBIF Togo qui a adressé ses vifs remerciements aux participants pour leur disponibilité et leur sérieux au cours des trois jours de travaux. Il a salué les résultats obtenus au cours de cet atelier et les a exhorté à maintenir l'effort et l'endurance pour réussir la



UNION EUROPÉENNE



3^{ème} partie de l'atelier qui est celui de la finalisation des drafts obtenus en vue de la publication effective des Data Papers. Il a également rappelé que le Nœud dispose de ressources financières pour aider au paiement des frais de publication des articles acceptés.

- Il a ensuite remercié le formateur Dr PANDO Francisco qui a accepté faire le déplacement du Togo malgré ses occupations au sein de son institution et a souhaité que cette collaboration dont les premiers fruits seront la publication des Data Papers soit poursuivie dans d'autres secteurs notamment dans celui du Jardin botanique.
- Le Node Manager GBIF Togo a enfin remercié les partenaires (Ministère de l'Enseignement supérieur et de la Recherche, Ministère de l'Environnement et des ressources forestières, ONG, UL (Département de Zoologie et Botanique) que sont les sectoriels qui favorisent le déroulement des activités du Projet BID au Togo.

CONCLUSION

Cet atelier au vu des participants, organisateurs et du formateur a atteint ses objectifs. En effet, ces trois jours de travaux ont permis à l'ensemble des participants présents à cet atelier de s'outiller sur la publication des data papers. Des draft de data papers issus des données individuelles ont été obtenus pour un travail de finition et des publications attendues

Fait à Lomé, le jeudi 14 décembre 2017

Le Rapporteur

Wiyaou GNAMA



UNION EUROPÉENNE



ANNEXES

Concept and plan for a Workshop on “Data Paper publication” within the context of the GBIF’s BID Project “Strengthening the biodiversity stakeholders network in Togo (BID-AF2015-0004-NAC)”

Prepared by Francisco Pando (Real Jardín Botánico-CSIC, Spain)

Trainer: Francisco Pando

Training Language: English

Contents

Objectives	7
Timeline	7
Participants	8
Requirements for a good data paper	8
Methods/facilities	9
Preliminary agenda.....	9

Objectives

Biodiversity data are often not easily accessible or visible. Among the obstacles to its publication is the lack of incentives for those gathering primary biodiversity data, among them, scientists and researchers. In accordance with the strategy of GBIF and some scientific publishers, this workshop proposes the publication of data articles (data papers) as a mechanism to give academic recognition to the effort made by data providers.

The specific objective of this workshop is that at the end of it, each participant has prepared and sent to a journal its own "data paper" based on their own datasets, using the GBIF's IPT (Integrated Publishing Tool).

A more general objective is to increase the capacity of the Togo's GBIF Node and network in data publication

Timeline

- 15 Nov: Concept doc released
- Workshop concept and requirements for a good data paper



21 -30 Nov. selection of participants

12-14 Dec. workshop

18 Dec.- 28 Feb. 2018. Helpdesk for data paper readying

Participants

This workshop is intended for researchers in Togo who:

- Already have a dataset suitable to be uploaded to the GBIF data network, to be described and with possibilities to be accepted as a "Data paper" in a recognized scientific journal.
- Have some proficiency in data handling (i.e. being able to sort, filter, find unique values in a table) with the tool of their election (preferably one of the following: MS-Access, SQL, OpenRefine)

Participants will be selected among candidates on these criteria and others indicated by the Togo's GBIF Node.

See next paragraph for what constitutes a "good dataset" in the context of this workshop

Requirements for a good data paper

The key concept when publishing a data paper is the potential the dataset has for re-usability.

What reusability means for a dataset:

- Provides a coherent view of a biodiversity question
 - All data records are provided under a common, current, recognized, documented taxonomy
 - Dataset constitutes a fair representation of the data known for the area, taxonomic group and topic which is the object of the dataset
- Provides a consistent and fairly complete view of a biodiversity question. A question may involve:
 - A geographic area that is biologically meaningful
 - A well-defined taxonomic group
 - A well-defined topic
- Data gathering and genesis methodology can be assigned and described for data records
- Accessible:
 - It is available from a recognized community repository (e.g. GBIF)
 - Data are provided under an accepted community standard
- The more potential users can be identified, the better. Examples of what may make a dataset "attractive" follow:
 - It has been prepared by an authority/ expert in the taxonomic group it comprises
 - It is large, compressive or covers an extensive territory
 - It is interested to a wide community (e.g. invasive or threatened species, species of economic importance)
 - It shows spatio-temporal patterns
 - Datasets has been analyzed in other (relevant) publications



Then the data paper must :
Provide rich metadata in an accepted standard

Methods/facilities

A google drive will be created by the host/organizer and shared with participants to act as a repository of preliminary datasets and other relevant materials for the workshop

An IPT in "Test mode" must be available to participants and instructor during the workshop to provide a platform to work with during the hands-on modules of the workshop

An Internet connection

A projector (beamer)

Participants must bring (or be provided with) laptops /PCs

Preliminary agenda

Day 1

09:00 - 09:30

1. Presentation.
 - 1.1. Welcome.
 - 1.2. Presentation of participants.
 - 1.3. Presentation of the course.

09:30 - 10:30

2. Introduction.
 - 2.1. The Biodiversity Information Landscape and the GBIF Role.
 - 2.2. Importance of publishing data through the GBIF.
 - 2.3. Open Science and Open Data.

10:30 - 11:00. Coffee break

11:00 - 12:30

- 2.4. Flow of data publishing through the GBIF.
3. Licenses.
4. Data standardization: Darwin Core, Darwin Core Archive.
 - 4.1 Darwin Core Data Standard: Terms, Concepts, and Examples.
 - 4.2 Darwin Core Archive.

12:30 - 14:00. Lunch break

14:00 - 15:30

- 4.3 Preparing a set of data according to the Darwin Core data standard (**practice**).

15:30- 16:00. Coffee break

16:00 - 17:00

5. Data Quality, Quality Control Concepts and data cleaning.
 - 5.1. Taxonomic data.
 - 5.2. Geographic data.
 - 5.3. People and entities, dates.
 - 5.4. Sensitive data.

Day 2

09:00 - 10:30

- 5.3 Data quality and cleaning (**practice**).
 - 5.3.1 [OpenRefine](#) / [DarwinTest](#): Demo and Examples.
 - 5.3.2 Data Cleanup Practice with OpenRefine/DarwinTest.
 - 5.3.3 Geographic Data Quality Control.



10:30 - 11:00. Coffee break

11:00 - 12:30

6. A data paper (article)

6.1 What is (and is not) a data paper.

6.2 Structure.

6.3 Examples.

7. Integrated Publishing Toolkit (IPT).

7.1 IPT in detail.

12:30 - 14:00. Lunch break

14:00 - 15:30

7.2 A guided walk through the IPT (**practice**).

15:30 - 16:00. Coffee break

16:00 - 17:00

7.3 Participant working with their own data and IPT (**practice**).

Day 3

09:00 - 10:30

7.3 Participants working with their own data and IPT (**practice**). Cont.

10:30 - 11:00. Coffee break

11:00 - 12:30

8. From IPT to data paper.

8.1 Write a data paper.

8.1.1 What sections does IPT do?

8.1.2 What sections must be incorporated manually?

8.2 Publishers and editors.

12:30 - 14:00. Lunch break

14:00 - 15:30

8.3 Examples of articles.

8.4 Participants completing paper sections not covered by IPT (**practice**).

8.5 Final steps and submission of data paper (**practice**).

9. Review, questions, clarifications.

9.1 Questions & Clarifications.

9.2 Evaluation and review.

15:30 - 16:00. Coffee break

16:00 - 17:00

10. Course closure.



UNION EUROPÉENNE



Annexes 2.

Introduire un document généré



UNION EUROPÉENNE

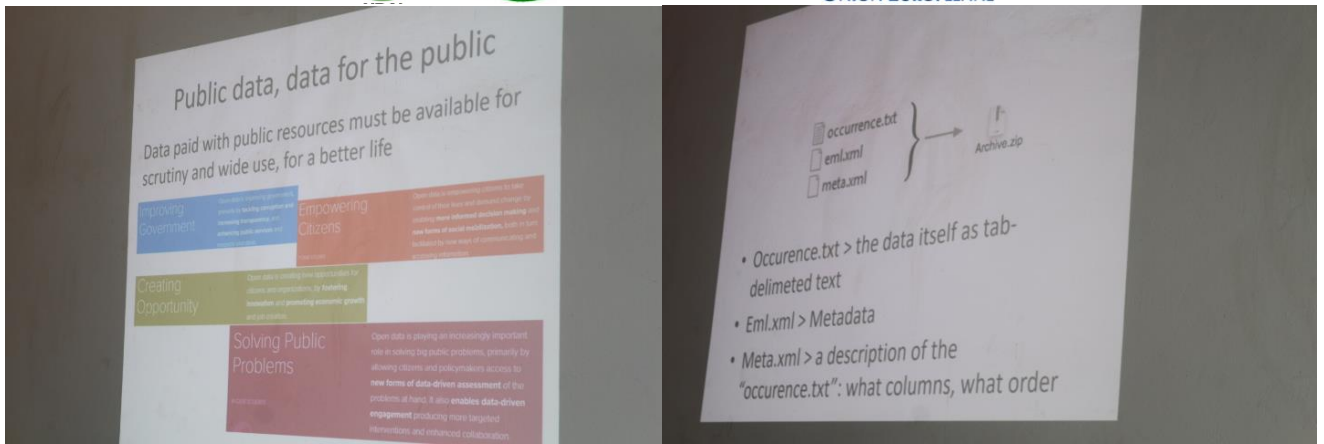


Annexes 3: Quelques photos





UNION EUROPÉENNE





Annexe 4: Liste et contacts des participants

**PROJET BID AU TOGO
DATA PAPER WORKSHOP
UNIVERSITE DE LOME, 12-14 DECEMBRE 2017**

LISTE ET CONTACTS DES PARTICIPANTS

N°	Nom et prénoms	Email address	Institution	Téléphone
1	AKPENE Koudjo	akpene.koudjo@gmail.com	WASCAL/UL	90038866
2	ALASSANI Stephane	stephane.shiny1995@gmail.com	FDS/UL	93169640
3	ASSOU Délagnon	patricedelagnon@gmail.com	ONG Agbo-Zégué	99772611
4	AWITAZI Tchagou	rodolphelmd@yahoo.fr	DRF/MERF	91862839
5	GNAMA Marc	marcgnama@gmail.com	DRF/MERF	90365806
6	GOMINA Mondjonnesso	gominadjo@gmail.com	Zoologie/UL	91856508
7	Mme AHONON Selom	ahononson@gmail.com	ITRA/DS	90176158
8	Mme TETE-BENISSAN Amivi	colette.gassou@gmail.com	FDS/UL	90038402
9	PALANGA Venant	vpalanga@gmail.com	MESR	90171260
10	POLORIGNI Botolisam	botolisam@yahoo.fr	MUHCV	90048139
11	RADJI Raoufou Pierre	pradji@hotmail.com	FDS/UL GBIF TOGO	90045114
12	SODJNOU Komlan	fidelesodjinou@hotmail.fr	FDS/UL	91930623
13	ZOGBEMA Jervis	zogbemajervis@gmail.com	MESR	90072202
14	PANDO Francisco	pando@rjb.csic.es	Jardin Botanique Royal Espagne	Formateur