



Compte-rendu du comité de pilotage GBIF France 17 juin 2014

Salle Jean Dorst, MNHN

Président de séance : Thomas Grenon, excusé, remplacé par Michel Guiraud

Rédaction

Anne-Sophie Archambeau (GBIF France)

Participants

Nom prénom	Structure
Evelyne Bégaud	Institut Pasteur
Thomas Changeux	IRD
Capucine Crosnier	MEDDE/PEM4
Michel Guiraud	MNHN
Yanick Lebeau	MEDDE/SINP
Chloé Martin	UMS BBES
Marc Pignal	MNHN/e-Recolnat
Laurent Poncet	SPN
Jean-François Silvain	FRB/IRD
Simon Tillier	MESR
Michèle Tixier-Boichard	INRA
Régine Vignes-Lebbe	UPMC/GBIF France
Eric Chenin	IRD/GBIF France
Anne-Sophie Archambeau	GBIF France

Objectif de la réunion

Comité de pilotage du GBIF France

Au-delà du présent CR, pour plus de détails sur les réalisations du GBIF et du GBIF France entre Janvier 2013 et Juin 2014, nous vous invitons à consulter le Rapport d'Activités, disponible sur le site www.gbif.fr à l'adresse <http://www.gbif.fr/page/ressources/rapports-comptes-rendus>.

Michel Guiraud ouvre la séance. Il rappelle l'ordre du jour et demande si des points sont à ajouter ou modifier : l'ordre du jour est accepté.



Après le rappel des missions du GBIF France, le bilan est présenté par Anne-Sophie Archambeau

Résultats du GBIF et du GBIF France 2013-mi 2014

Recensement et connexion des données au portail GBIF

Le GBIF connecte 450 millions de données en juin 2014. La France est le 7ème pays contributeur du portail GBIF, avec 16,49 millions de données connectées dont 9 millions sont géoréférencées ; les spécimens sont plutôt bien représentés avec plus de 6,5 millions.

En France, 111 jeux de données sont actuellement connectés, les deux plus importants fournisseurs étant le MNHN pour les spécimens et le SPN (Service du Patrimoine Naturel) pour les observations. De nouveaux jeux de données ont été connectés cette année, ceux-ci étaient de taille modeste mais ont contribué à la diversification des jeux de données connectés. La liste taxonomique TAXREF a également été indexée cette année.

Les données de l'ONF (Office National des Forêts) et de la FCBN (Fédération des Conservatoires Botaniques Nationaux) devraient bientôt être connectées via l'INPN (Inventaire National du Patrimoine Naturel).

Suite à l'appel à proposition lancé en juin 2012 pour des projets de numérisation de collections ou de création de listes taxonomiques, plus de 54.000 nouvelles données ont pu être recueillies et connectées au portail international du GBIF. (Le budget total était de 42 K€ et 11 projets avaient été retenus)

Qualité des données

Le GBIF poursuit un travail de fond sur les données et notamment sur la qualité des données : détecter et traiter les doublons, détecter les données erronées (taxonomie, géoréférencement) et les signaler aux détenteurs, etc.

Afin d'aider les fournisseurs et utilisateurs de données, le GBIF Espagne a identifié et décrit différents outils d'aide au nettoyage des données et les a mis en ligne dans le « Biodiversity data quality hub » (<http://www.gbif.es/BDQ.php>) (en anglais).

Utilisation du portail GBIF

L'utilisation des données accessibles via le portail GBIF est toujours difficile à mesurer.

La consultation du portail et le téléchargement de données sont en forte croissance, mais l'utilisation correspondante, qu'elle soit liée à la recherche ou à l'action et à la prise de décision, n'est pas facile à identifier.



Le Mexique et les Etats-Unis sont quasi systématiquement en tête de la liste des 10 pays qui téléchargent le plus de données (établie tous les mois), avec notamment CONABIO, l'outil de la politique mexicaine en environnement qui est très actif. On retrouve régulièrement la France dans ce top 10.

Le nombre des publications scientifiques qui citent le portail GBIF comme moyen d'accéder aux données utilisées augmente (67 en 2008, 229 en 2012, 249 en 2013) mais il reste faible, notamment comparé aux grands portails thématiques.

Une explication pourrait être qu'il est utilisé comme référence pour la découverte de données et pour des opérations de recherche préliminaires -comme par exemple la vérification de la présence d'une espèce dans une zone géographique donnée- ; mais que lorsque l'opération de recherche s'approfondit, elle fait appel à des données spécifiques, plus faciles à trouver sur des portails thématiques spécialisés, et que ce sont ces sources qui sont alors citées dans les publications, disciplinaires par nature. Ces sources peuvent d'ailleurs avoir été identifiées via le GBIF, mais cela n'est pas précisé dans les publications.

Afin de favoriser la citation des données, le GBIF va attribuer, courant 2014, un DOI à chaque jeu de donnée et toujours un UUID pour chaque occurrence. Cela devrait faciliter la traçabilité.

Remarque : Il faudrait essayer de comparer la différence d'utilisation d'une base de donnée avant ou après sa connexion au GBIF.

Le portail national

Le développement du portail national avait avancé en 2013 (moteur de recherche et métadonnées) mais des changements vont intervenir. En effet, suite au développement par l'Australie de son portail national : Atlas of Living Australia (ALA), plusieurs nœuds GBIF ont émis le souhait de pouvoir l'implémenter dans leur pays. Cependant il était au départ très difficile à transposer car peu documenté et plutôt fermé. Après discussion avec ALA, un workshop aura lieu fin juillet à Canberra avec toute l'équipe de ALA et une dizaine de développeurs de différents points nodaux GBIF dont la France et l'Espagne afin d'étudier la possibilité de le réutiliser. A l'issue de ce workshop, la décision de déployer le portail national à partir des modules ALA devrait être actée.

Cela permettrait de constituer des portails nationaux basés sur des technologies semblables et de créer un réseau de développeurs au niveau international. Cela contribuerait à une meilleure réutilisation des outils développés par différents nœuds (et notamment ceux développés par le GBIF France) et enfin, permettrait de lutter contre les pertes de compétences et de connaissances liées aux changements de personnel (fréquents du fait des CDD).

Le portail national permettra des analyses plus fines sur les données françaises connectées au GBIF que le portail du GBIF international ne le permet. Il offrira plus de souplesse pour filtrer, visualiser, naviguer et exploiter les données.

Nous veillerons à ce qu'il y ait une bonne articulation avec l'INPN : des liens seront par exemple créés vers les distributions d'espèces présentes sur les territoires français. Pour



rappel, une partie des données françaises connectées au GBIF ne concernent pas les territoires français et ne sont pas visibles sur le site de l'INPN.

Formations

Le GBIF France a participé à l'organisation du GB20 training course « building stronger GBIF nodes » qui a eu lieu en octobre 2013, avant le conseil d'administration. 57 participants de 33 pays ont suivi cette formation. Le GBIF France a traduit tous les documents en français et Anne-Sophie Archambeau était formatrice sur 2 modules.

Afin de développer l'utilisation des données au niveau national, une formation a été organisée en partenariat avec le projet BioVeL les 24-25 mars 2014 au MNHN. Le but : aperçu des différentes façons de publier, accéder et utiliser les données sur la biodiversité, via le réseau GBIF et les outils BioVeL. BioVeL est un laboratoire virtuel qui met à disposition des « workflows » pour traiter les données de la recherche sur la biodiversité.

<http://www.gbif.fr/content/formation-sur-la-qualite-lutilisation-et-la-publication-de-donnees-sur-la-biodiversite-24-25>

Cette formation a été filmée et enregistrée et sera intégrée dans la future plate-forme d'e-learning qui sera mise en place à la fin du mentoring (voir plus loin).

Communication

Le GBIF France a fait un effort particulier cette année pour développer et diversifier ses actions de communication.

L'équipe a participé à de nombreuses manifestations en 2013 et 2014 : congrès scientifiques, journées sur l'open data, ...

Le GBIF France a également profité du lancement officiel du nouveau portail GBIF international en octobre 2013 pour organiser une journée-événement le 9 décembre 2013 à l'auditorium de l'Aquarium Tropical de la Porte Dorée à Paris.

Pour cette occasion, et en collaboration avec la société Zip Shebam, nous avons travaillé sur le logo GBIF France et le design du nouveau site web et nous avons fait appel à la société Jordenen pour la fabrication de goodies GBIF France : sacs coton, stylos, post-its, blocs-notes.

Le but de cette journée était de présenter les nouvelles fonctionnalités du portail international ainsi que les activités du point nodal et de faire intervenir nos partenaires.

De nombreux exemples d'utilisation des données sont présents sur le nouveau portail et des pages spécifiques par pays ont été créées pour mieux documenter l'effort réalisé par les pays participants.

150 personnes ont participé à cette journée et le retour a été très positif.

<http://www.gbif.fr/content/presentation-du-nouveau-portail-de-donnees-et-du-reseau-gbif-9-decembre-2013>



Le site web du GBIF France a également été remanié et est en cours de finalisation. Un soin tout particulier sera apporté aux pages partenaires afin de mieux les valoriser.

Interactions au niveau national

E-ReColNat (ANR-11-INBS-0004 e-ReColNat) a pour objectif de réunir l'ensemble des données des collections françaises d'histoire naturelle sur une même plateforme informatique.

E-ReColNat est conduit par le MNHN et rattaché à la Direction des Collections. Il se consacre au montage et au scan des spécimens de l'herbier de Montpellier, au scan de spécimens de zoologie et de paléontologie du MNHN, et à la saisie des données relatives aux spécimens à partir de leurs images, dans un esprit de science participative. E-ReColNat se consacre aussi à la visibilité des collections à l'international, et **intègre en son sein le GBIF France**.

Le GBIF France apporte son expertise sur :

- les outils de connexion
- l'interopérabilité des bases de données de biodiversité
- l'utilisation des standards de données internationaux utilisés en biodiversité informatique (Darwin core, ABCD...).

Il contribue principalement aux work packages 2 (Coordination informatique des données) et 7 (Valorisation). Depuis le lancement du projet, le GBIF France a fortement contribué à la communication du projet et au débogage de la version 1 du site des Herbonautes.

Le GBIF France poursuit également ses collaborations avec les différents programmes et institutions au niveau national :

- **INPN et SINP** : participation aux groupes de travail - plate-forme SINP
- **FRB** : Ecoscope, Biovel, CESAB
- **RBDD** : intervention aux journées RBDD
- **UMS BBES, OCIM, FCBN,...**

et poursuivra ces actions à l'avenir.

Le GBIF France a notamment réalisé une étude sur les identifiants uniques (état de l'art, préconisation) pour le SINP.

Il suit également les informations sur les standards au niveau international (TDWG).

Interaction au niveau européen

Le GBIF France participe chaque année aux réunions des nœuds européens: Joensuu, Finlande en 2013 et Bruxelles, Belgique en 2014.

Le réseau des nœuds GBIF européen est solide et actif. Il existe une forte volonté de monter des projets européens ensemble et notamment dans le cadre d'H2020.

Le **projet « OpenUp »** (open-up.eu) est un projet européen (ICT PSP fourth call for proposals 2010 Best Practice Network) porté par Walter Berendsohn de la Freie Universitaet de Berlin.

> Accès libre et gratuit aux données de la biodiversité

Son objectif: contribuer au site web culturel européen EUROPEANA (<http://www.europeana.eu>) en mettant à disposition des ressources multimédia sur la biodiversité via les infrastructures et les standards GBIF.

Le projet a démarré en Mars 2011 pour une durée de 3 ans et le consortium était constitué de 23 institutions scientifiques (dont le MNHN) provenant de 12 pays européens différents (<http://open-up.eu/node/8>). Le projet s'est terminé fin février 2014 avec une très bonne évaluation de la Commission Européenne.

Le MNHN a connecté 513.321 images provenant de différentes collections du MNHN et le GBIF France a rempli son contrat. Lors du congrès de clôture du projet en février 2014 à Bratislava, Slovaquie, (<http://open-up.eu/node/1267>), le MNHN et le GBIF France ont été félicités pour le travail accompli.

Le projet Open-Up ! est terminé mais Europeana est un site web pérenne qui à vocation à rester en ligne et auquel il sera toujours possible de fournir de nouvelles données multimédia car l'infrastructure de connexion des données reste en place.

Le projet **EUBON** (www.eubon.eu) propose une approche innovante en matière d'intégration de systèmes d'information, des données de terrain aux données de télédétection, pour répondre aux besoins des politiques. Objectif : renforcement de la composante européenne pour une mise en œuvre globale de GEO BON et soutien à l'IPBES. C'est un projet de 5 ans géré par les Allemands (30 partenaires - 18 pays).

Le MNHN n'étant pas partenaire officiel du projet, le **GBIF France demande** au comité de pilotage son accord pour qu'il **devienne Associate Partner d'EUBON** puisque le MNHN est l'entité légale qui peut s'engager.

La demande est acceptée.

Mentoring France Espagne Portugal

Les points nodaux français, espagnols et portugais ont répondu à un appel à mentoring du Secrétariat du GBIF. Huit mille euros nous ont été accordé pour travailler sur les thèmes suivants : formation & e-learning, data paper, data visualisation et crowdsourcing.

Trois meetings sont prévus:

Madrid : 21-23 janvier 2014

focus sur data quality et training & e-learning

Lisbon : 28-30 Avril 2014.

focus sur les data papers et la visualisation des données.

Paris : 14-16 October 2014

focus sur la data visualisation et le crowdsourcing

Dans le cadre de ce mentoring, le GBIF France travaille notamment sur le développement de nouvelles fonctionnalités pour les herbonautes et aidera l'Espagne et le Portugal à monter un projet de crowdsourcing tandis que la France et le Portugal auront accès à la plateforme de e-learning espagnole.



L'expérience acquise lors de ce mentoring sera également bénéfique pour d'autres points nodaux. En effet, les sites de crowdsourcing et les plate-formes de e-learning mis en place existeront dans les trois langues et pourront être consultés par tous les pays francophones, hispanophones et lusophones.

La partie visualisation des données se fera en lien avec les développements réalisés par le groupe de développeurs créé autour du workshop ALA.

Fonctionnement et budget du point nodal

En dehors des coordinateurs du GBIF France, Eric Chenin et Régine Vignes-Lebbe qui sont à temps partiel et respectivement à l'IRD et à l'UPMC, l'équipe qui travaille à temps complet n'est constituée que de CDD : une node manager (IRD en détachement MNHN), 2 ingénieurs informatiques et une ingénieure données. Cela engendre un turn-over important de l'équipe et fragilise le suivi des actions.

Budget :

Le total des dépenses en 2013 s'est monté à 313 193 euros (similaire à 2012).

Pour 2014, la prévision est de 324 659 euros.

Le bilan budgétaire est validé mais il est peu lisible. Il faudra veiller à mieux le présenter en séparant bien les ressources budgétaires (e-recolnat, open-up, mentoring, ...) et en respectant mieux l'intitulé des personnels.

Il faudrait demander une prise en charge des missions par les structures demandeuses lorsque le GBIF France intervient en province.

Soutien au Sud

L'appui entamé dans le cadre du programme SEP CEPDEC sera poursuivi. Il le sera dans le cadre de projets de mentoring, financés par le Secrétariat du GBIF ; et dans le cadre de la nouvelle phase de Sud Expert Plantes, qui devrait démarrer fin 2014 – début 2015. Cette nouvelle phase prévoit de poursuivre le soutien apporté aux pays du Sud (22 pays francophones d'Afrique, de l'Océan indien et d'Asie du Sud-Est) pour que ceux-ci participent efficacement au GBIF ; il prévoit aussi d'étendre aux données d'abondance les formats de données pris en compte par le portail GBIF. La collaboration avec CEPDEC (programme du Secrétariat) devrait également se poursuivre.

La poursuite du soutien par SEP à la participation des pays du Sud au GBIF, est notamment motivé par l'importance du manque de données sur la zone inter-tropicale, qui est patent sur le portail GBIF : en particulier le manque de données récentes ; la zone étant assez bien couverte dans les périodes anciennes, grâce aux données des collections des Muséums et Herbiers du Nord.

L'appui au Sud qu'apportera le GBIF France concernera essentiellement la sensibilisation des pays non encore adhérents, la formation et l'assistance technique auprès des points nodaux



nationaux, et l'hébergement des données pour les points nodaux et les détenteurs qui le demandent.

Perspectives fin 2014-2015

Portail national :

Participation au workshop de Canberra puis implémentation du portail si la décision est confirmée. Bien vérifier l'articulation et la compatibilité des outils ALA et INPN.
Le faire tester par le groupe de travail utilisateurs.

Connexion :

Continuer l'aide aux éditeurs de données.

Data paper : organisation d'une formation en 2015.

Quand la plate-forme SINP existera, les identifiants uniques UUID attribués par le SINP seront récupérés par le GBIF.

Utilisation du portail et des données

Développer l'analyse et la promotion de l'utilisation, meilleure visibilité des statistiques de téléchargement sur les fournisseurs de données, synthèse des travaux réalisés par Wala Adraa.

Mentoring

Suivi des projets du mentoring avec l'Espagne et le Portugal : mise en place de la plateforme de e-learning française et organisation du troisième workshop à Paris en octobre.

Collaboration avec les autres dispositifs

Identifier les tâches et outils d'intérêt commun

Organiser les flux

Organiser la collaboration entre les équipes

Interactions au niveau européen et international

Signature du MoU de EUBON

Suivi des projets et des réunions

Appui au Sud

Mentoring

Contribution à SEP/DD

Hébergement de données

Divers

Evelyne Bégau : Congrès international sur les collections européennes. Pasteur : Mai-Juin 2015 Paris ; présentation GBIF ; réseau des Instituts Pasteur : CI, RCA, Sénégal, Madagascar

Question : comment sont présentés les COM, sous quel pays ? (Nouvelle Calédonie) ?



Global Biodiversity Information Facility
Système Mondial d'Information sur la Biodiversité

> Accès libre et gratuit aux données de la biodiversité

