Table ronde n°2

Les bases de données sont utilisées par la communauté scientifique dans le cadre de ses recherches. Cette activité se nourrit des bases de données mais les nourrit aussi. Quel contrôle d'accès aux données (droit des producteurs)? Existe-t-il des modèles de contrôle efficaces ? Quel mode de citation de la base et du jeu de données d'origine (traçabilité, respect des producteurs). Quel équilibre entre charge de gestion et d'enrichissement du système?

AllEnvi
Alliance nationale de recherche pour l'Environnement

« Infrastructures, systèmes d'observation et services d'information environnementale et climatique »

MUSÉUM NATIONAL D'HISTOIRE NATURELLE

Les données accessibles dans le domaine environnemental.

Quels accélérateurs ?

François ROBIDA

Géosciences pour une Terre durable
brgm

Des réalisations opérationnelles (infrastructures de données et de **services**)

> **Infoterre** et web services - 12 millions de visites par an sur nos sites web et nos services http://infoterre.brgm.fr/

> **OneGeology** : 117 pays pour produire la carte géologique du monde de façon distribuée http://portal.onegeology.org/

> **Géocatalogue :** le catalogue national d'INSPIRE (8000 métadonnées) http://www.geocatalogue.fr/

> **Carmen** : outil de publication de données environnementales : 180 producteurs, 8000 services web OGC, 10000 visites/jour http://www.carmencarto.fr

brgm

Des mots clés

- **Interopérabilité** = « L' aptitude à coopérer » (donc la volonté…)

- **Standards** : rien sans standards – données géospatiales (yc capteurs) on n'a pas le choix en dehors des couches sémantiques (ontologies)

  OGC® Open Geospatial Consortium, Inc. <GeoSciML/>

  (Toutes les infras sont bâties sur les mêmes composants)

- **Mutualisation** : ne pas réinventer, partager les outils, les infrastructures, les savoir faire...

- Dans le respect de la **subsidiarité** : chacun reste maître de ses données

- L'enjeu principal : **publier - publier - publier …**
  et cataloguer pour réutiliser

Géosciences pour une Terre durable brgm

Un proposition d'action vers un objectif commun pour Allenvi

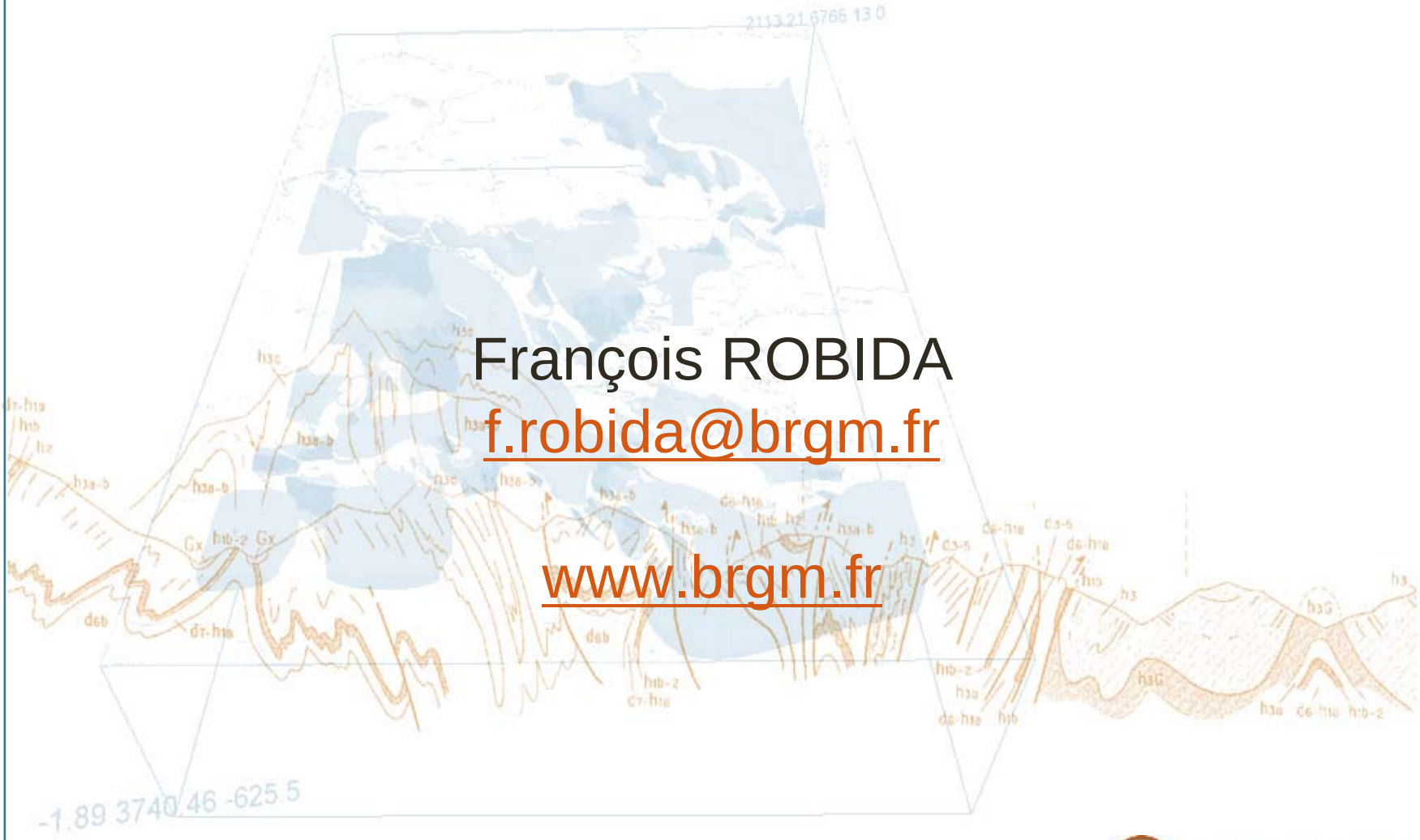
> Rendre visible la masse d'observatoires d'ALLENVI au niveau national en construisant « l'observatoire français de l'environnement »

→ Constituer le pôle OPEN DATA de l'environnement

> Plusieurs types d'actions :

- Formations
- Fédérer les catalogues (moissonnage ou saisie en direct)
- Mutualiser les ressources (logiciels de catalogage, publication,…) et infrastructures (serveurs, « cloud privé»),…
- Mutualiser les efforts sur les technologies et les standards
- Peser au niveau européen et international (GEO/GEOSS, OGC,…)
- Une appropriation commune des nouveaux outils (brokering approach, linked data,…)

Géosciences pour une Terre durable
brgm



François ROBIDA

f.robida@brgm.fr

www.brgm.fr

Géosciences pour une Terre durable

brgm

Paris, 21 mai 2014

Global Biodiversity Information System (GBIF)

gbif@gbif.fr

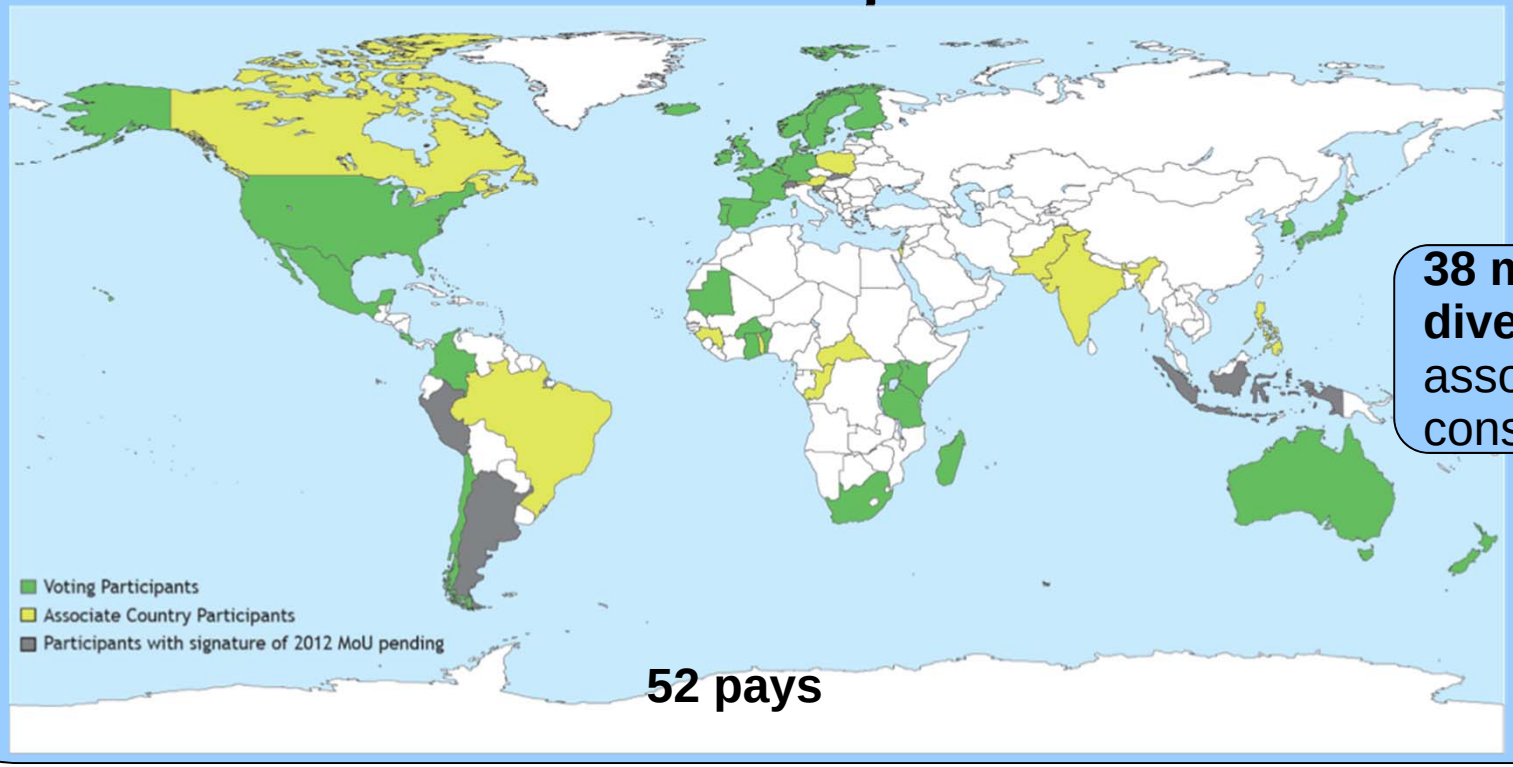
www.gbif.org / www.gbif.fr

Organisation du GBIF

Governing Board
Comité exécutif
Comité pour le budget
Comité des noeuds participants
Comité scientifique

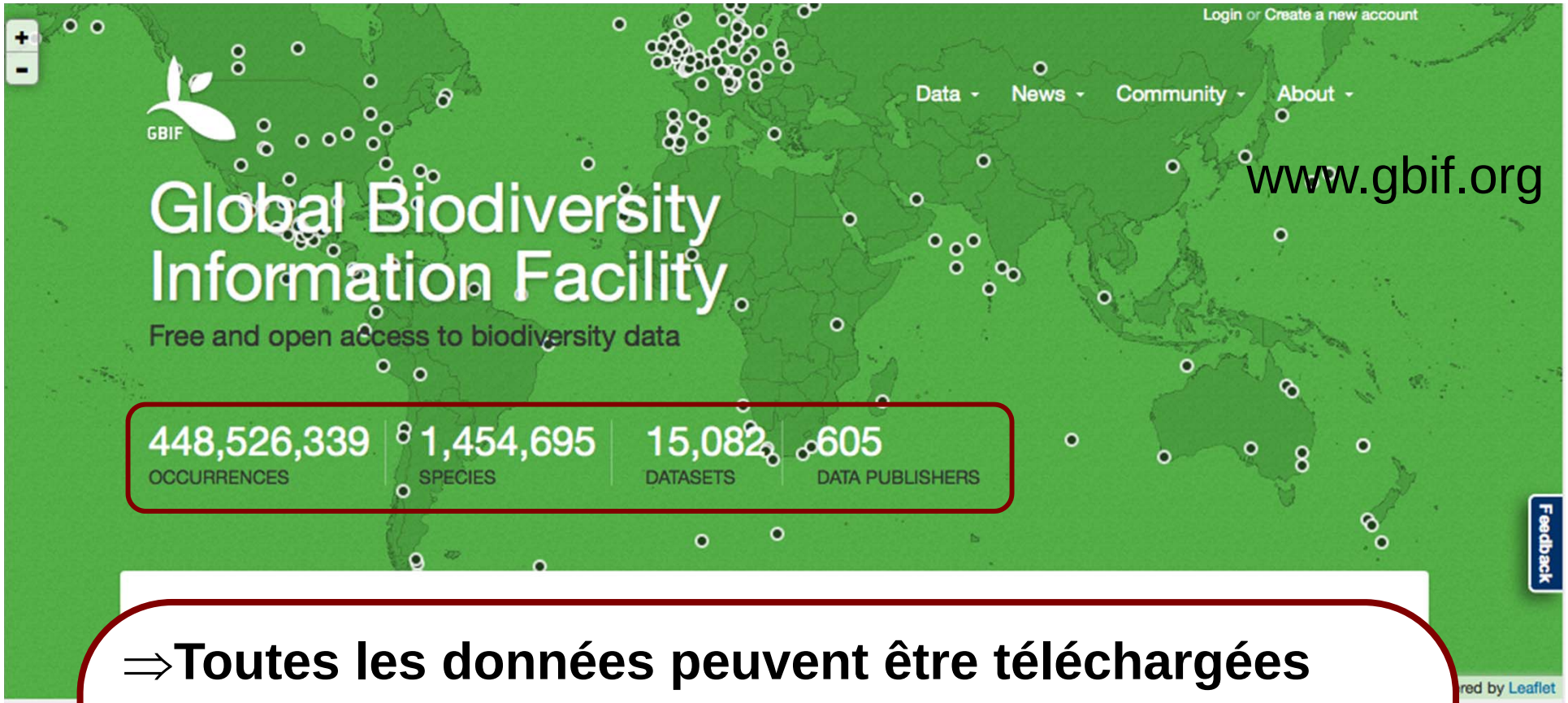
Secrétariat
Basé à
Copenhague

90 Participants



52 pays

38 membres divers : associations, consortiums, …



⇒**Toutes les données peuvent être téléchargées** librement et gratuitement et **réutilisées par d'autres portails** (différents **API**: registry, maps, taxonomy…).

⇒ **Tous les outils** développés par le GBIF sont **gratuits** et **open source** (IPT pour publication, DwC validator …) et utilisent **des standards internationaux** (Darwin Core, ABCD, EML…).

Login or Create a new account

GBIF.ORG
Free and open access to biodiversity data

Data News Community About

http://data.rbge.org.uk/herb/E00647233

Specimen of Prumnopitys andina (Poepp. ex Endl.) de Laub. recorded on 13 janv. 2013

from Royal Botanic Garden Edinburgh Herbarium (E) dataset

Information

Location

LOCALITY

Forest along the north side of the Río Diguillan adjacent to Reserva Nacional Ñuble;, Chile

-71,49, -36,969

ELEVATION

930m ± 0m

GEOGRAPHIC CLASSIFICATION

Chile > Región VIII [Biobío] > Provincia de Ñuble

En cours de développement:

Mise en place d'un modèle de citation basé sur les DOI.

⇒**Attribution d'un DOI à chaque jeu de données** et toujours un UUID à chaque occurrence

⇒ toutes les citations référenceront les jeux de données utilisés à travers leur DOI.

Citation and licensing

The content of the "Dataset citation provided by the publisher" depends on the metadata supplied by the publisher. In some cases this may be incomplete. A standard default form for citing is provided as an alternative. We are in transition towards providing more consistent citation text for all datasets.

DATASET CITATION PROVIDED BY PUBLISHER

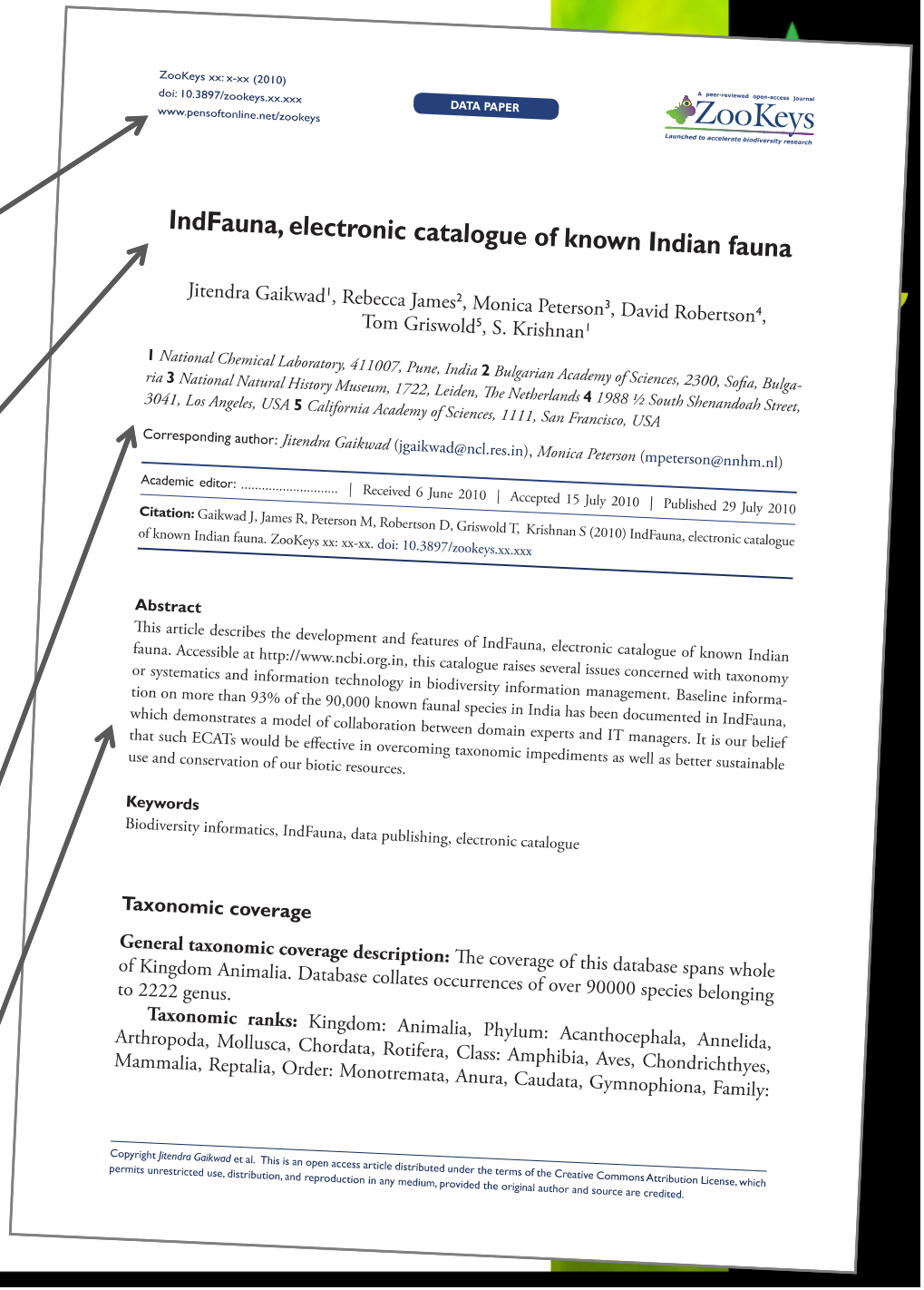
Royal Botanic Garden Edinburgh. (2014) Preserved Collections Database (E)

DEFAULT CITATION

Royal Botanic Garden Edinburgh: Royal Botanic Garden Edinburgh Herbarium (E), 2014-05-10.
Accessed via http://www.gbif.org/occurrence/890166540 on 2014-05-19

Data Paper

- DOI : indexation et citation
  - indexés par Web of Knowledge (ISI), PubMedCentral, Scopus, Zoological Record, Google Scholar, CAB Abstracts, DOAJ, EBSCO.
- *Promouvoir et faire connaître les données*
- Reconnaissance des éditeurs de données via une publication scientifique
- Décrit les données sous forme structurée et lisible par un humain



ZooKeys xx: x-xx (2010)
doi: 10.3897/zookeys.xx.xxx
www.pensoftonline.net/zookeys

DATA PAPER

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

IndFauna, electronic catalogue of known Indian fauna

Jitendra Gaikwad[1], Rebecca James[2], Monica Peterson[3], David Robertson[4], Tom Griswold[5], S. Krishnan[1]

1 *National Chemical Laboratory, 411007, Pune, India* **2** *Bulgarian Academy of Sciences, 2300, Sofia, Bulgaria* **3** *National Natural History Museum, 1722, Leiden, The Netherlands* **4** *1988 ½ South Shenandoah Street, 3041, Los Angeles, USA* **5** *California Academy of Sciences, 1111, San Francisco, USA*

Corresponding author: *Jitendra Gaikwad* (jgaikwad@ncl.res.in), *Monica Peterson* (mpeterson@nnhm.nl)

Academic editor: ..... | Received 6 June 2010 | Accepted 15 July 2010 | Published 29 July 2010

Citation: Gaikwad J, James R, Peterson M, Robertson D, Griswold T, Krishnan S (2010) IndFauna, electronic catalogue of known Indian fauna. ZooKeys xx: xx-xx. doi: 10.3897/zookeys.xx.xxx

Abstract

This article describes the development and features of IndFauna, electronic catalogue of known Indian fauna. Accessible at http://www.ncbi.org.in, this catalogue raises several issues concerned with taxonomy or systematics and information technology in biodiversity information management. Baseline information on more than 93% of the 90,000 known faunal species in India has been documented in IndFauna, which demonstrates a model of collaboration between domain experts and IT managers. It is our belief that such ECATs would be effective in overcoming taxonomic impediments as well as better sustainable use and conservation of our biotic resources.

Keywords

Biodiversity informatics, IndFauna, data publishing, electronic catalogue

Taxonomic coverage

General taxonomic coverage description: The coverage of this database spans whole of Kingdom Animalia. Database collates occurrences of over 90000 species belonging to 2222 genus.

Taxonomic ranks: Kingdom: Animalia, Phylum: Acanthocephala, Annelida, Arthropoda, Mollusca, Chordata, Rotifera, Class: Amphibia, Aves, Chondrichthyes, Mammalia, Reptalia, Order: Monotremata, Anura, Caudata, Gymnophiona, Family:

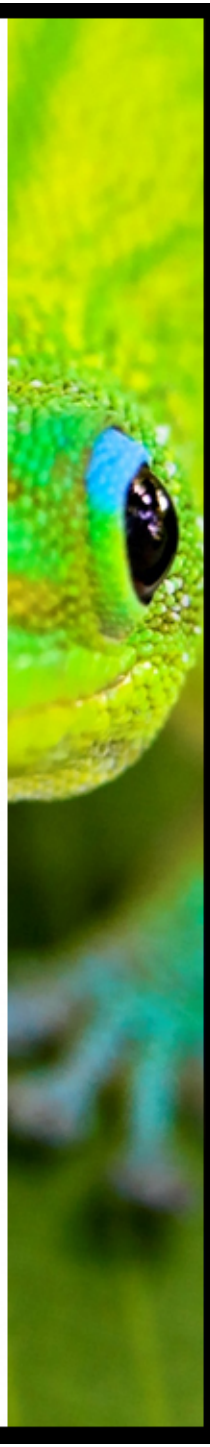
Copyright *Jitendra Gaikwad* et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Propriété Intellectuelle

•Une **consultation GBIF** sur les licences de données a été lancée en aout 2013 et a débouché sur une **proposition ouverte à commentaires jusqu'au 14 juin 2014** : http://www.gbif.fr/?p=1821

- Licence CC0 ou équivalent,
- modification du processus de publication des données : flag sur les jeux de données dont les fournisseurs veulent empêcher l'utilisation à des fins commerciales.
- « Normes de publication et d'usage des données », établies par la communauté GBIF (par exemple les guides de bonnes pratiques concernant la citation des données, ainsi qu'une redéfinition de l'usage commercial des données).

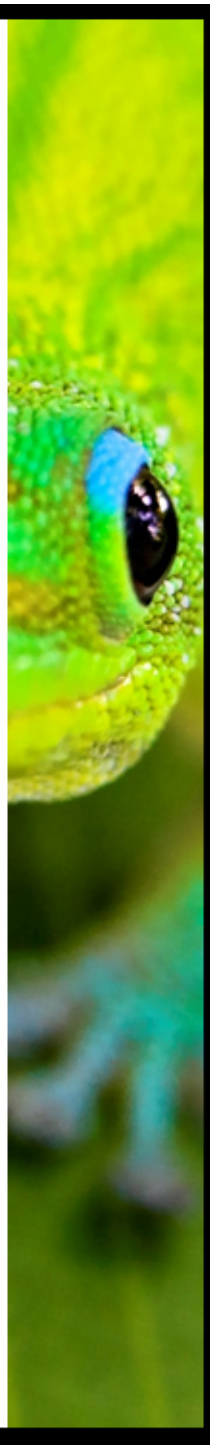
•Mise en place d'un **processus de lecture des licences de données directement par les machines** et recommandations sur les usages pour toutes les données mobilisées dans le GBIF. prévu fin 2014



Interactions au niveau national

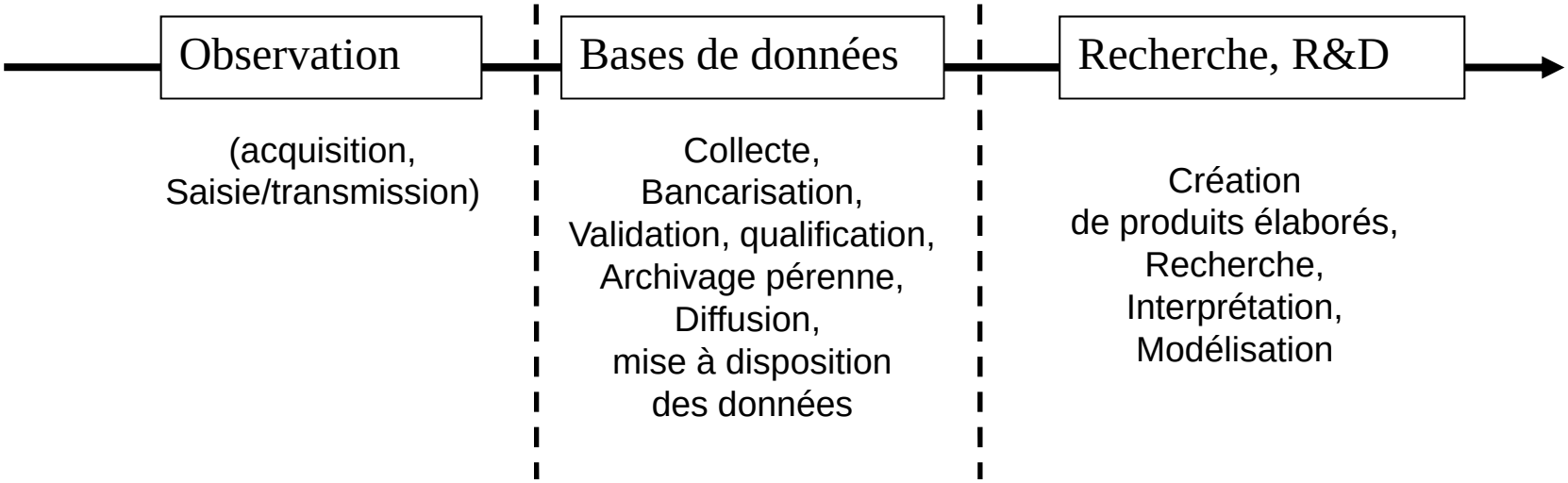
Le GBIF France travaille actuellement en collaboration avec différents programmes et institutions au niveau national : e-Recolnat, INPN, FRB, ECOSCOPE, SINP, UMS BBEES, organismes, etc...., et poursuivra ces actions à l'avenir. (harmonisation des flux de données, interopérabilité…)

Pour la connexion des données, si un interlocuteur particulier existe sur un niveau ou une thématique donnée, le GBIF s'adresse à ce point d'accès là en priorité.



GBIF
www.gbif.fr

IFREMER : Les bases de données marines



Phase « observation » :

Principaux producteurs de données

- TGIR UMS FOF : Flotte Océanographique Française
- TGIR Euro-Argo (et projet inter-organisme Coriolis)
- SIH (Système d'Information Halieutique)
- Réseaux d'observation pour la surveillance de l'environnement et du littoral
- Missions satellite intéressant l'océan

IFREMER : Les bases de données marines

Phase « bases de données » :

- Organisation par filières
- Données et méta-données
- Les méta-données répondent notamment aux questions : Quoi? Où? Quand? Comment?
- Traçabilité : nécessité scientifique et reconnaissance du producteur

Phase « Recherche, R&D » :

- Nombreuses et anciennes collaborations sur les sciences marines aux niveaux national et international
- La mise en commun des données est une nécessité scientifique et technique (contraintes/volumes parfois importants)

IFREMER : Les bases de données marines

Caractéristiques des données marines :

- variations temporelles
- géo-référencement
- océan = domaine partagé
- open data avant l'heure

L'activité « bases de données » à Ifremer

- ~35 ETP
- ~2,5 M€ de budget annuel
- organisation en unité de service
- la gestion des droits fait partie intégrante de la gestion de données

IFREMER : Les bases de données marines

UNESCO : IODE (Intergovernmental Oceanographic Data Exchange), OBIS

UE : Seadatanet, Emodnet, MyOcean, Geo-Seas,…

National : Pôle Océan (et les 3 autres pôles de données d'observation de la terre)

Cadre réglementaire : INSPIRE, DCSMM, DCE, DCF, code minier

Principes de traçabilité, de contrôle d'accès, de citation souvent établis dans des projets en coopération (PPCM des contraintes de chacun)

Exemples : DOI sur jeux de données, fédération d'identités

Les données accessibles dans le domaine environnemental. Quels freins ? MNHN, 21 mai 2014

Table ronde 2

La politique Irstea en matière de gestion des données de la recherche

Pour mieux affirmer ses missions, le Cemagref devient Irstea

www.irstea.fr

Emmanuelle Jannès-Ober,
Pôle recherche & innovation, Déléguée IST, DP2VIST
emmanuelle.jannes-ober@irstea.fr

Etat des **lieux** : une mosaïque de pratiques en héritage

- Très faible culture de la valorisation commerciale : essor fragile depuis 2009 avec le développement du libre accès (Irstea est signataire de la déclaration de Berlin depuis 2006, loi sur l'accès aux données publiques, H2020) ;

- Sauf exception, faible culture de la sécurité des données et de la propriété : un cadrage juridique et technique depuis 2009 avec la création de la direction de la valorisation et du transfert (travail de fond en matière d'acculturation des équipes et de cadrage technique) ; et le recrutement d'un juriste spécialisé en propriété intellectuelle. Une exception : base avalanches (EPA/CLPA) car en délégation de service public. En matière de capitalisation, des pratiques disciplinaires (hydrologie) ;

- Pas de réponse technique interne : des bases de données scientifiques en cours de référencement par la DSI et d'analyse des contenus (un catalogue des bases scientifiques en cours : difficulté pour mobiliser les scientifiques sur l'existant, appui des personnels IST prévu) ;

*= > un **inventaire** partiel : difficulté pour déterminer parfois ce qui appartient à Irstea et à des tiers (cédé gratuitement ou acheté) ; difficulté pour déterminer ce qui est éligible à une valorisation commerciale et ce qui peut être diffusé en libre accès*

=> urgence de mettre en place une politique d'établissement !

La politique de gestion des données de la recherche en marche

1. Une **politique globale qui intègre l'ensemble des modalités de valorisation** (économique et non économique) liées au système qualité d'Irstea : définition et déclinaison dans un guide pratique destiné aux scientifiques.

2. Le guide pratique pointe vers une quinzaine de fiches pratiques qui renvoient à des procédures Irstea intégrées aux processus qualités (Gérer les données issus des travaux de recherche ; Valoriser des travaux de recherche).

3. Un projet d'établissement collectif et transversal 2014 : toutes les directions du pôle recherche & innovation + direction en charge des systèmes d'information et la direction en charge des questions juridiques. Portage : direction de la stratégie de la recherche et de l'évaluation, DVT et DP2VIST.

4. Ce qui est défini : documenter ses données, formats de fichiers - standards, réaliser un plan de management des données, rendre visible ses données, sécuriser, archivage, dépôt APP, contractualisation, protection juridique d'une BDD, transfert de données à caractère personnel, générer des revenues à partir d'un BDD, recours face aux usages frauduleux, open data et H2020.

5. Ce qu'il reste à faire : valider la politique Irstea, établir la répartition des tâches entre les acteurs internes ; réaliser le plan de communication interne. Priorité 2015 DP2VIST (Dir. Prospective, veille et valorisation non économique de l'IST).

La politique de gestion des données de la recherche en marche

1. Quel contrôle d'accès aux données (droit des producteurs)? Existe-t-il des modèles de contrôle efficaces ?
   - Irstea veut garantir les droits de ses partenaires (prévention par la documentation des données, la sensibilisation des scientifiques). Distinction entre données brutes et produits (dont BDD)
   - Pour garantir les droits d'Irstea : contractualisation ; processus de veille
   - Le modèle Freemimum pour les bases de données

2. Quel mode de citation de la base et du jeu de données d'origine (traçabilité, respects des producteurs) ?
   - Obligation de citation (métadonnées des bases diffusées ; remerciement dans les publications)

3. Quel équilibre entre charge de gestion et d'enrichissement du système?
   - Au préalable, définir le rôle de chaque acteur (entre recherche et fonctions d'appui) : enrichissement du système (scientifiques et IST) ; gestion (UR, Valorisation économique – DVT, DSI, Dir. juridique, IST) ;
   - Définir le processus et les procédures : mobilisation des acteurs sur les nouveaux projets, dès la réponse à un appel d'offre et au démarrage des projets pour parvenir à un équilibre, garantir les droits, permettre un accès pérenne (payant ou gratuit suivant les cas).