GBIF
www.gbif.org

AP-BON Science Symposium, 5 Dec 2011, Tokyo, Japan

Biodiversity data standards, infrastructures and interoperability for GEO BON – the GBIF contribution

Dr Éamonn Ó Tuama
Senior Programme Officer, Inventory, Discovery, Access (IDA)
Global Biodiversity Information Facility (GBIF)

United Nations Decade on Biodiversity

Outline

- GBIF informatics architecture & network
- Use of standards to enable interoperability
- Uses of GBIF mediated data
- GBIF and GEO BON

GBIF
www.gbif.org

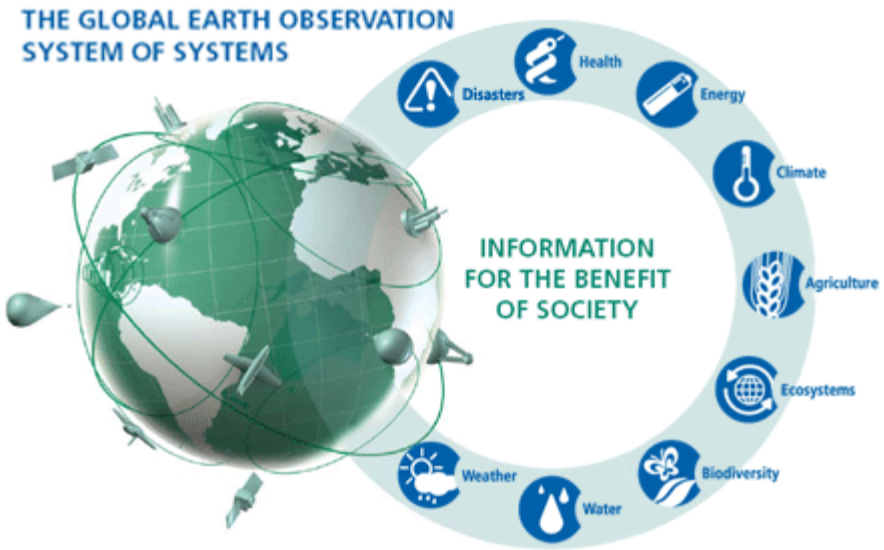
2011-2020
United Nations Decade on Biodiversity

GBIF network / architecture

About GBIF

The mission of the Global Biodiversity Information Facility (GBIF) is to facilitate **free and open access to biodiversity data** worldwide via the Internet to underpin **scientific research**, **conservation** and **sustainable development**.

GBIF can support several GEOSS Societal Benefit Areas



GBIF
www.gbif.org

United Nations Decade on Biodiversity

Components of data sharing

1. The ***data*** – the 'content' of the system;
2. The ***system*** – infrastructure, tools and services ('pipes & plumbing') to allow data to flow – enables data publishing, discovery, access
3. The ***people*** – users, data publishers, legislators, enablers, communities of practice, etc

All three are necessary, none is sufficient of itself….

GBIF
www.gbif.org

United Nations Decade on Biodiversity

Key Data Sharing Challenges

- Data management;
- Data documentation (metadata);
- Data discovery (catalogues);
- Data access (web services, use tracking);
- Data exchange standards and protocols;
- Data types – what is needed?
- Data quality – fitness-for-use?
- Data volumes – how much is enough?
- Data security - open access vs sensitive data;
- Benefits/incentives for data sharing?
- Data ownership, IPR;
- Data attribution (authorship, recognition)

GBIF
www.gbif.org

2011-2020
United Nations Decade on Biodiversity

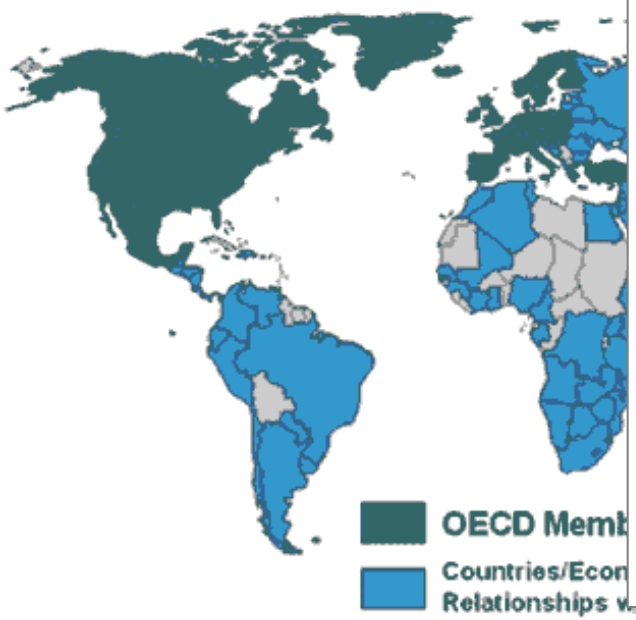
The **OECD origin…**

OECD
ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

GBIF
www.gbif.org

OECD Global Science Forum recommendation (1999):

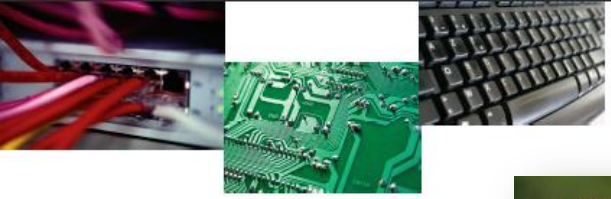
"Establish and support a distributed system of interlinked and interoperable modules (databases, software and networking tools, search engines, analytical algorithms, etc.) that together will form a **Global Biodiversity Information Facility (GBIF)**"

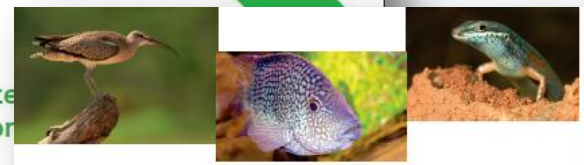


"This facility will enable users to navigate and put to use vast quantities of biodiversity information, thereby:

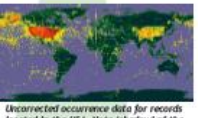
- advancing ***scientific research…***
- serving *the* ***economic…***
- providing *a basis from which our* ***knowledge of the natural world*** *can grow rapidly…"*

United Nations Decade on Biodiversity





The GBIF Data Portal: a gate… biodiversity infor…



Uncorrected occurrence data for records located in the USA. Note 'shadow' of the USA over Asia, due to incorrect longitude entries.

Occurrences for the USA as seen from mid-2011, following processing improvements

Occurrences for the USA as seen from late 2011, following further improvements. Note coastal zones are now included.

Background

The GBIF Data Portal has provided discovery and acc… the content shared through GBIF since its launch in … resource accessible through the portal is primary sp… data such as specimens in museums, observations in… culture strains.

Issues

The primary critiques received of the portal relate t… content and the time lag between data being added … the network, and becoming visible on the portal. Ot… identified for improvement were the backbone taxo… organize the content, which had become disorderly … rapidly in number; and the geographic and temporal … of records (see maps in left column).

Solutions – improvements to the portal in 2011

During 2011, GBIF has developed new processing wo… include the following:

- reduced processing time for data from 3-4 days to … enabling more frequent rollovers (publication cycl… therefore shorter delay between entering data an… on portal;
- a complete regeneration of the taxonomic backbo… up-to-date taxonomic catalogues;
- enhancements of the registry (see poster on GBIF … support better modelling of the GBIF network stru… and interactions. This paves the way for better at… data owners and service providers; and
- extended data interpretation routines and plausib… e.g. in the areas of geolocation, date interpretati… depth, etc. (for illustration of some results, see m… column).

The solutions were developed to run on the Hadoop … efficient parallel processing environment enabling t… to grow with future data volumes.

The GBIF Registry: improving discovery and access

Components and Interactions

Integrated Publishing Toolkit (IPT)
Whenever anybody installs an IPT instance in their institution, they can register it to the GBIF Network. Any resources being hosted by that IPT instance can be published as well. This is possible because the IPT interacts directly with the GBIF Registry's Web Service API.

Harvesting and Indexing Toolkit (HIT)
The HIT is the tool responsible for harvesting and indexing biodiversity data from access points throughout the GBIF network. It has a dependency on the Registry's Web Service API, used to discover all available access points.

Other GBIF participant systems
There are a number of external GBIF participant systems making use of the Registry's Web Service API for their own purposes. The API documentation is simple and clear, so it is quite straightforward to make use of the API with minimal technical knowledge.

Refactoring to improve
The GBIF Registry codebase has been improving through time to meet the needs expressed by the GBIF community. Currently all the code is open for anybody to make use of and hopefully improve. It is an ongoing development, as requirements are always being added.

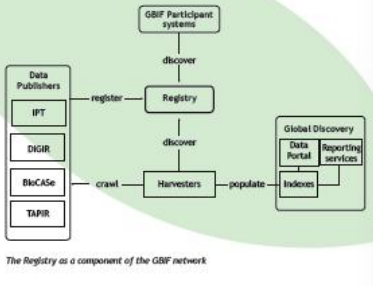
One of the major challenges for existing biodiversity informatics infrastructure is to enable users to improve substantially their ability to discover and access relevant biodiversity information and data resources.

The GBIF Registry

The GBIF Registry is a critical component of the GBIF network. Its main purpose is to hold the information about the entities that constitute the whole GBIF network, as it stores metadata on Participant nodes, institutions and resources along with their contact information. In order to enable discovery of this information the GBIF Registry exposes a Web Service Application Programming Interface (API) to allow third party tools (e.g. harvesters) to discover and access biodiversity data. This API also offers the possibility to publish/update/delete resources on the network, using the proper authorization credentials.

Monitoring the GBIF Network

As the status of the network is ever-changing, the GBIF Registry needs to be able to detect these changes and take appropriate action (e.g. new contact info for a resource). A newly-developed monitoring service keeps track of 1) the availability (online/offline status) of the access points throughout the network, 2) changes in number of records shared throughout time and 3) changes in metadata information. This service runs periodically to keep the Registry's information as up to date as possible.



The Registry as a component of the GBIF network

Building a distributed metadata system for the GBIF network

The metadata system supports open exchange protocols. In particular, it supports the Open Archives Initiative for Metadata Harvesting (OAI-PMH[2]), offering the possibility of integration with other catalogue systems, e.g. Metacat[3] and Geonetwork[4]. The GBIF portal, in turn, can forward the aggregated metadata from its network to other clearinghouses such as the EuroGEOSS broker[5] thereby facilitating discovery of biodiversity resources on complementary networks. Work is currently in progress to support those GBIF Participants wishing to connect their national metadata catalogues to the GBIF network.

Large, distributed networks such as GBIF's bring together many publishers and consumers of data. To guide consumers in discovering data that is appropriate for their needs and fit for use, all datasets when published should be accompanied by metadata, in a standard format, that describe critical aspects of the data such as sampling procedures and methods, data quality, provenance, ownership, data format, access and intellectual property rights. Once generated, metadata are typically stored in online catalogues (databases) that can be browsed and searched.

Metadata are a central component in an expanding GBIF network, and it is essential that information about biodiversity datasets is well organized.

To meet this need better, GBIF is implementing a metadata system for its network that provides unified access to all participating metadata catalogues[1]. The principal components of the system are i) a central metadata catalogue holding copies of all metadata published on the network, ii) one or more participating external metadata catalogues and iii) a set of protocols and data exchange standards to allow flow of metadata in the network. These components and the technologies chosen for their implementation, are described in Figs. 1 and 2.

Figure 1. The GBIF metadata network

The central metadata catalogue is being integrated in the improved GBIF data portal (Fig. 3), which allows a user to search for resources across the three main data types available (occurrences, checklists, metadata). Encouraging the publication of high quality, complete metadata via the distributed catalogue system will enable potential end users to discover easily which datasets are available and, critically, to evaluate the appropriateness of such datasets for particular purposes.

[1] http://links.gbif.org/gbif_metadata_catalogue_specification.pdf
[2] http://www.openarchives.org
[3] http://knb.ecoinformatics.org/index.jsp
[4] http://geonetwork-opensource.org/
[5] http://www.eurogeoss.eu/broker/

GBIF
www.gbif.org
…free and open access to biodiversity data

Figure 2. The system architecture includes three main functional components: a set of online accessible metadata catalogues, a metadata harvester, and the GBIF central metadata catalogue. Software choices for implementation are as indicated

Figure 3. GBIF data portal: integration of metadata

About GBIF

The Global Biodiversity Information Facility (GBIF) was established by governments in 2001 to encourage free and open access to biodiversity data, via the Internet. Through a global network of national and thematic nodes, and a Secretariat based in Copenhagen, Denmark, GBIF promotes and facilitates the mobilization, access, discovery and use of information about the occurrence of organisms over time and across the planet.

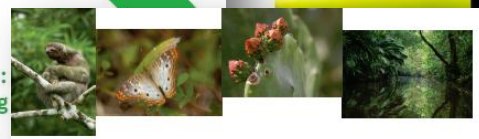
Vision - A world in which biodiversity information is freely and universally available for science, society, and a sustainable future.

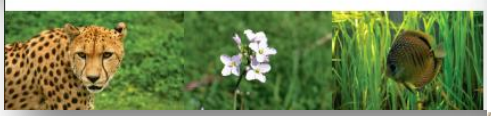
Mission - To be the foremost global resource for biodiversity information, and engender smart solutions for environmental and human well-being.

www.gbif.org

2011-2020 United Nations Decade on Biodiversity

http://www.gbif.org/communications/resources/posters/





GBIF
www.gbif.org

IPT: facilitating biodiversity c

Overview

The GBIF Integrated Publishing Tool
developed to facilitate efficient bi
Internet. The IPT can manage three
occurrence data, (ii) taxonomic ch
metadata. Through it, users are ab
local databases, upload existing file
delimited formats, and access cent
standardized controlled vocabulari
following the International standar
(DwC-A) and the Ecological Modelir

The IPT installed by TaxIDF (http://www.taxidf.or.cr/ipt/) already holds 27 datasets describing the biodiversity of the country.

After considerable community feed
revision in early 2011. The new ver
and use, more efficient and more s
production of Data Papers (see sep
which represent a tangible incentiv
publishing.

IPT Uptake

Since the release of IPT v. 2, 27 ins
with GBIF, connecting a total of 10
260,000 checklist records to the ne
approximately 25 per cent of the d
network this year.

The uptake and deployment of the
IPT helpdesk and training experts,
Site. Many of the group members w
Workshop on the GBIF IPT, held in C
experts are now actively supportin
and training others on its use world

The Canadian organization Canadensys has customized the IPT to match the style of its existing website (http://data.canadensys.net/ipt/)

GBIF

Simplified data pul

Spreadsheet Publishi

Spreadsheets are commonly used by biologists to manage biodiversity data. GBIF has developed a series of spreadsheet files that provide simple templates for data entry for species occurrence and species checklist data. When the complete spreadsheets are uploaded to a GBIF processor application, a publication-ready file is returned.

Darwin Core Archive Assi

The Darwin Core Archive Assistant is a service that allows database managers to produce a publication-ready data file without installing any local tools. They can describe their data files, and the service will produce the required data descriptor file that forms part of the data exchange standard format. This file is bundled with the biodiversity data files for distribution through the GBIF network.

The Data Paper:
an incentive for metadata publishing

The lack of incentives for data publishers is one of the key impediments to free and open access to biodiversity data. The Data Paper has been conceived as one possible mechanism to offer scholarly recognition for the effort and investment involved in authoring rich metadata, and publishing them as a citable academic paper.

A Data Paper is a searchable metadata document describing a particular online accessible dataset, or a group of datasets, and published in accordance to the standard academic practices. Its purposes are:

- to provide a citable journal publication that brings scholarly credit to data publishers;
- to describe the data in a structured human-readable form; and
- to bring the existence of the data to the attention of the scientific community.

Unique features of Data Papers include: (a) low technology and infrastructural overheads; (b) close links or interconnections between data publishing and scholarly publishing cycles; (c) an automated, push-button conversion tool exporting metadata to a manuscript and (d) minimal core metadata elements to reduce time required for authoring a metadata document.

Data Papers for biodiversity will use the GBIF Metadata Profile (GMP) to author the metadata document. The GBIF Integrated Publishing Toolkit (IPT) contains a user-friendly interface that makes authoring metadata easy.

Data flow

A list of IPT installations supporting authoring of the Data Paper is accessible at http://tools.gbif.org/data-paper-authoring/. Guidelines for authors and reviewers of the Data Paper are accessible at http://www.pensoft.net/J_FILES/Pensoft_Data_Publishing_Policies_and_Guidelines.pdf.

The Data Paper prototype was developed in collaboration with PenSoft Publishers (http://www.pensoft.net/index.php)

Data publishers will be credited through
(a) registering authorship in a scholarly publication,
(b) indexing and citation of Data Papers similar to research papers,
(c) tracing usage and citations of published data, and
(d) providing a persistent description of the published data resource over time.

GBIF
www.gbif.org
...free and open access to biodiversity data

Taxonomic name finder

Full-text documents represent a vast corpus of knowledge about species and are increasingly being made available online. Keyword search works well for identifying potentially useful information when a species name is known, but extracting a list of species to build an index from a document has been primarily a manual process.

TaxonFinder is a name-finding web service offered by GBIF. It offers standardized input and output formats that allow scientific name extraction to be embedded within project workflows. For example, TaxonTagger (see diagram) is a web application, that uses the service. A document is loaded and all scientific names are highlighted and extracted. This list can be passed to other GBIF name services, for example to provide a complete and up-to-date taxonomic classification for all found species names.

The TaxonTagger process

About GBIF

The Global Biodiversity Information Facility (GBIF) was established by governments in 2001 to encourage free and open access to biodiversity data, via the internet. Through a global network of national and thematic nodes, and a Secretariat based in Copenhagen, Denmark, GBIF promotes and facilitates the mobilisation, access, discovery and use of information about the occurrence of organisms over time and across the planet.

Vision - A world in which biodiversity information is freely and universally available for science, society, and a sustainable future.

Mission - To be the foremost global resource for biodiversity information, and engender smart solutions for environmental and human well-being.

www.gbif.org

2010

http://www.gbif.org/communications/resources/posters/

Comprehensive capacity building programme

http://www.gbif.org/communications/resources/posters/

Mobilising new sources of biodiversity data

August 2011 IAIA

Special Publication Series No. 7

Publishing EIA-Related Primary Biodiversity Data

PURPOSE

The overall purpose of this guide is to enable EIA practitioners, consultants and other interested and affected parties to discover, capture, manage and publish to common standards, the primary biodiversity data generated during environmental impact assessment processes.

It represents a summarized version of a more comprehensive guide (ISBN: 87-92929-35) that can be accessed at http://links.gbif.org/eia_biodiversity_data_publishing_guide_en_v1

ABOUT GBIF

The Global Biodiversity Information Facility (GBIF) was established by countries as a global mega-science initiative to address one of the great challenges of the 21st century – harnessing knowledge of the Earth's biological diversity. GBIF's mission is to be the foremost global resource for biodiversity information, and to engender smart solutions for environmental and human well-being. To achieve this mission, GBIF encourages a wide variety of biodiversity data holders, generators and users across the globe to discover and publish (make discoverable) data through the GBIF network. For more information, visit http://www.gbif.org

HOW TO CITE THIS PUBLICATION

Cadman, M., Chavan, V., King, N., Willoughby, S., Rajvanshi, A., Mathur, V., Roberts, R., and Hirsch, T. (2011). Publishing EIA-Related Primary Biodiversity Data: GBIF-IAIA Best Practice Guide. IAIA Special Publication Series No. 7. August 2011. Pp. 6. Accessible at http://www.iaia.org/publications/

COLLABORATORS

This best practice guide is an outcome of the GBIF led collaboration of the following institutions:

Global Biodiversity Information Facility
South African National Biodiversity Institute
Wildlife Institute of India
International Association for Impact Assessment

INTERNATIONAL ASSOCIATION for IMPACT ASSESSMENT

- *Headquarters*

1330 23rd Street South, Suite C
Fargo, ND 58103-3705 USA
Phone +1.701.297.7908
Fax +1.701.297.7917
info@iaia.org
www.iaia.org

Publishing EIA-Related Prima
Biodiversity Data:
GBIF-IAIA Best Practice Guid

International Best Practice Principles

"PUBLISHING" BIODIVERSITY DATA MAY BE DEFINED AS
BIODIVERSITY DATASETS PUBLICLY ACCESSIBLE IN A ST
FORMAT, VIA AN ONLINE ACCESS POINT (TYPICALLY A
URL). THIS ACCESS POINT IS RECORDED IN A REGISTR
GLOBAL BIODIVERSITY INFORMATION FACILITY (GBIF).
CAN ALSO BE DISCOVERED AND ACCESSED VIA THE G
(HTTP://DATA.GBIF.ORG).

Introduction

The issue

Primary biodiversity data is defined as "digital text or multimedia da
about the occurrence of an organism." Knowledge about the identity
isms forms the backbone of our understanding of the biological w
monitoring the state of natural ecosystems, for developing sound env
policies, and making ecologically sustainable development decisions
Assessment (EIA) provides opportunities for integrating biodiversity
but, for a variety of reasons, biodiversity has not always been give
consideration in EIAs (Rajvanshi *et al.*, 2007).

Ideally, ***biodiversity-inclusive*** EIA, which is promoted by the Conventi
should: (a) use biodiversity information to determine the biologica
of a site, and (b) generate new biodiversity records about the site.
sessments, EIA practitioners need access to verifiable biodiversity
usable form and that can be accessed using standardized protocols
has been no easy-to-use mechanism for discovering and accessing di
use in EIA, or for publishing the biodiversity data that EIA generate

This means that EIA-related biodiversity data is, generally, unavailal
EIAs, or for informing research programmes, environmental planni
This compromises the quality of the EIA, reduces the transparency
ultimately, the confidence that can be placed in decisions based on

The solution

Through the ***Global Biodiversity Information Facility*** (GBIF), digi
being made freely and openly available via the Internet for scientists
and the general public. GBIF provides a suite of standards and da
can be employed to discover and publish primary biodiversity data.
describes the tools, standards and infrastructure that are available
explains when and how they should be used. It represents a summ
comprehensive guide (ISBN: 87-92929-35) that can be accessed at
biodiversity_data_publishing_guide_en_v1. Sources of additional ass

Improving EIA practice: Best Practice Guide for publishing primary biodiversity data Version1.0

IAIA

GBIF
www.gbif.org

Improving EIA practice: Best Practice Guide for publishing primary biodiversity data

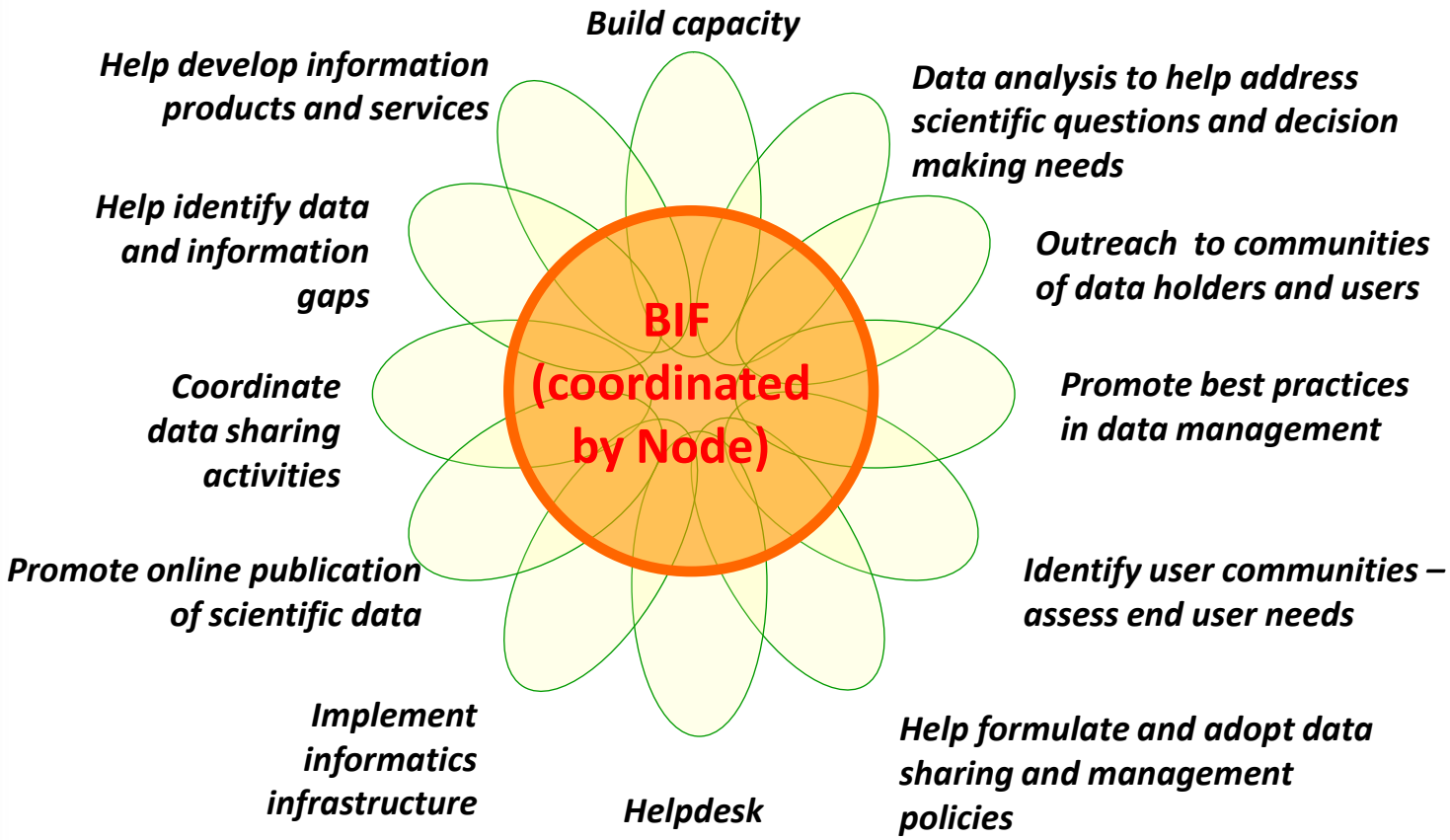
Version 1.0

August 2011

GBIF
www.gbif.org

United Nations Decade on Biodiversity

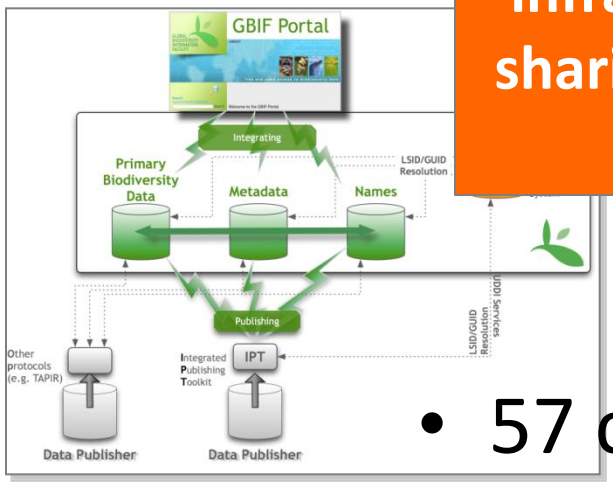
BIFs established by a country or organisation to serve multiple needs – env, agric, health, forestry, climate change, etc....



GBIF
www.gbif.org

United Nations Decade on Biodiversity

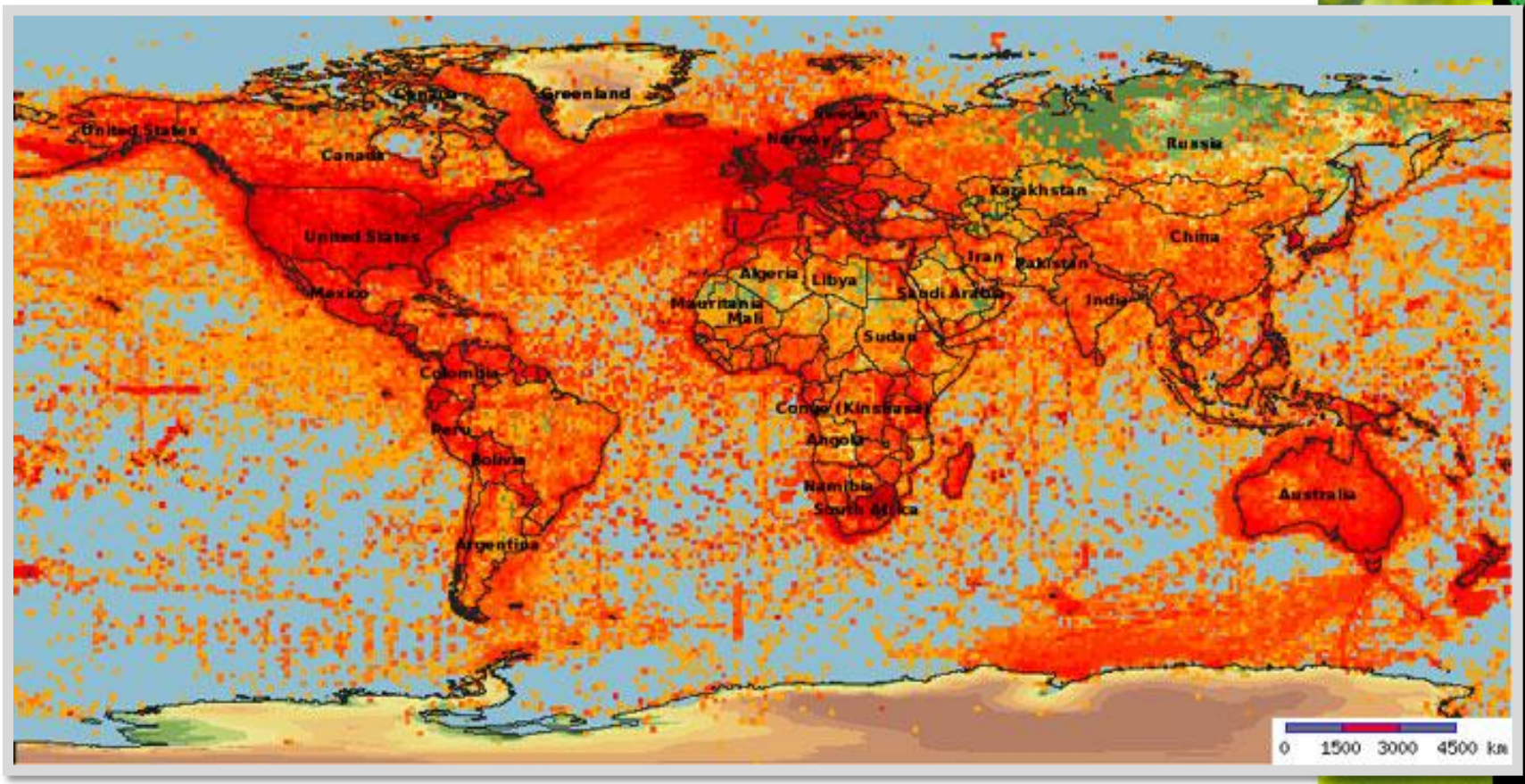
A global infrastructure for data exchange/sharing via a single portal



A distributed global infrastructure for the sharing of biodiversity data

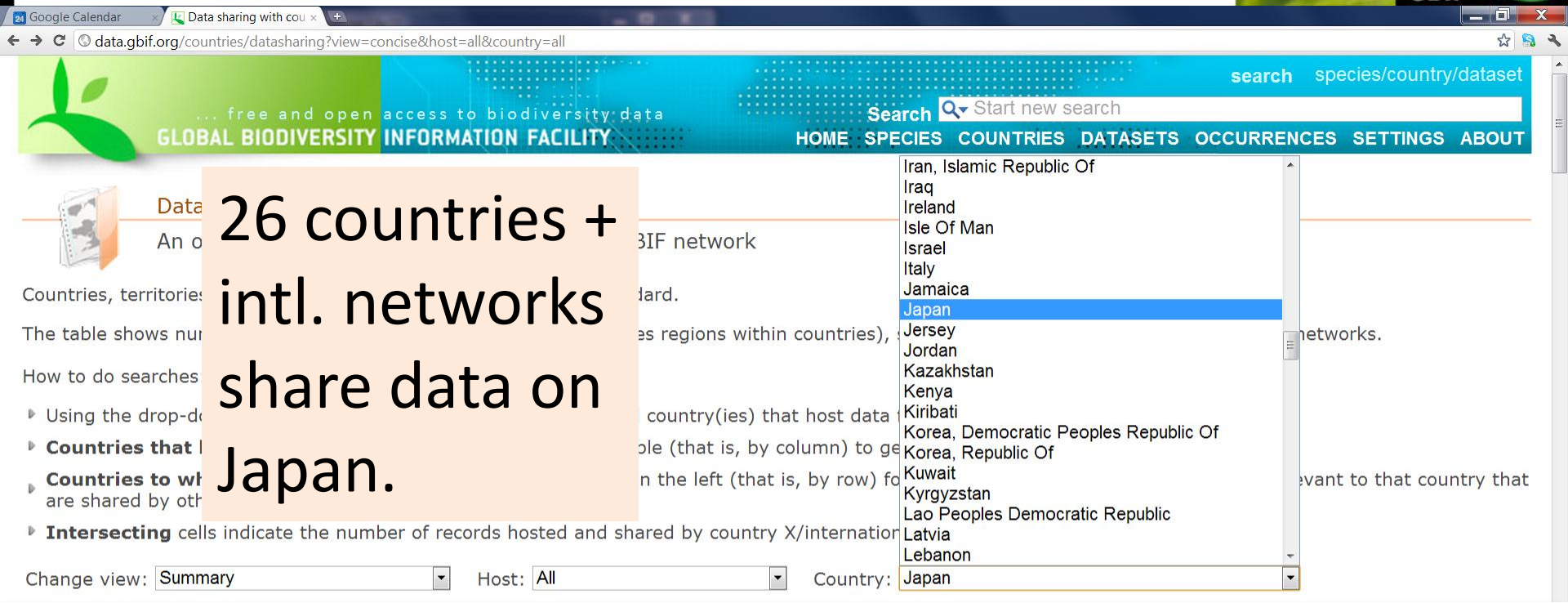
- 57 countries
- 45 organisations

Current GBIF-Network Data Coverage



Nov 2011: >**312** million occurrence records from >**10,000** datasets from >**300** publishers and spanning a wide range of geospatial, temporal and taxonomic coverages.

Universal open access…



26 countries + intl. networks share data on Japan.

Japan shares data on 198 countries.

Achieving interoperability

- Simplified data standards
  - Darwin Core Archives
- Improved data publication processes
  - Integrated Publishing Toolkit (IPT)
  - Spreadsheet-based publishing
- Improved integration
  - New taxonomic resources
  - Improved data processing

GBIF
www.gbif.org

2011-2020
United Nations Decade on Biodiversity

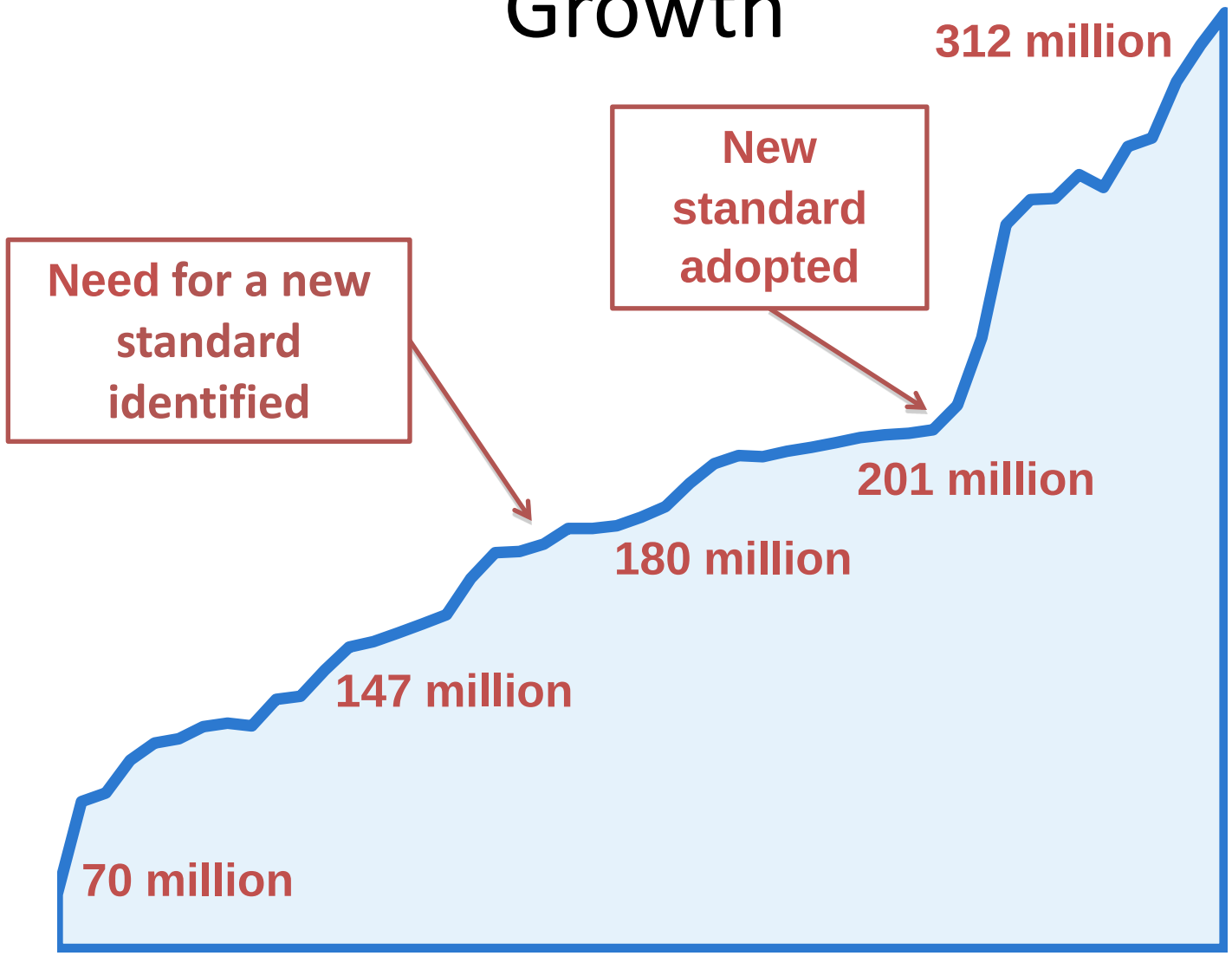
Darwin Core Archives (DwC-A)

A text-based solution to publishing biodiversity data

GBIF
www.gbif.org

2011-2020
United Nations Decade on Biodiversity

Growth



GBIF
www.gbif.org

United Nations Decade on Biodiversity

Promotion of Publishing Standards

http://www.gbif.org/orc/

> 2500 downloads English/French/Spanish

Taxonomic catalogues

GBIF taxonomic backbone is based on many authoratative catalogues:

- Catalogue of Life 2011
- Fauna Europeaa
- The National Center for Biotechnology Information (NCBI)
- The Integrated Taxonomic Information System (ITIS)
- **… and 40+ others**

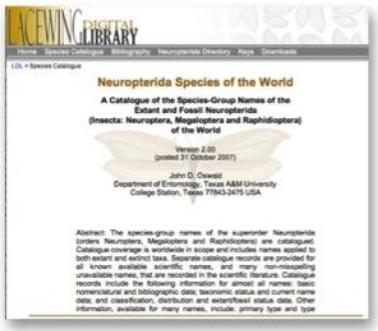
GBIF
www.gbif.org

United Nations Decade on Biodiversity

Enabling authoratative taxonomic data to be published through GBIF

> 50 new checklists

Taxonomic



Regional

Thematic

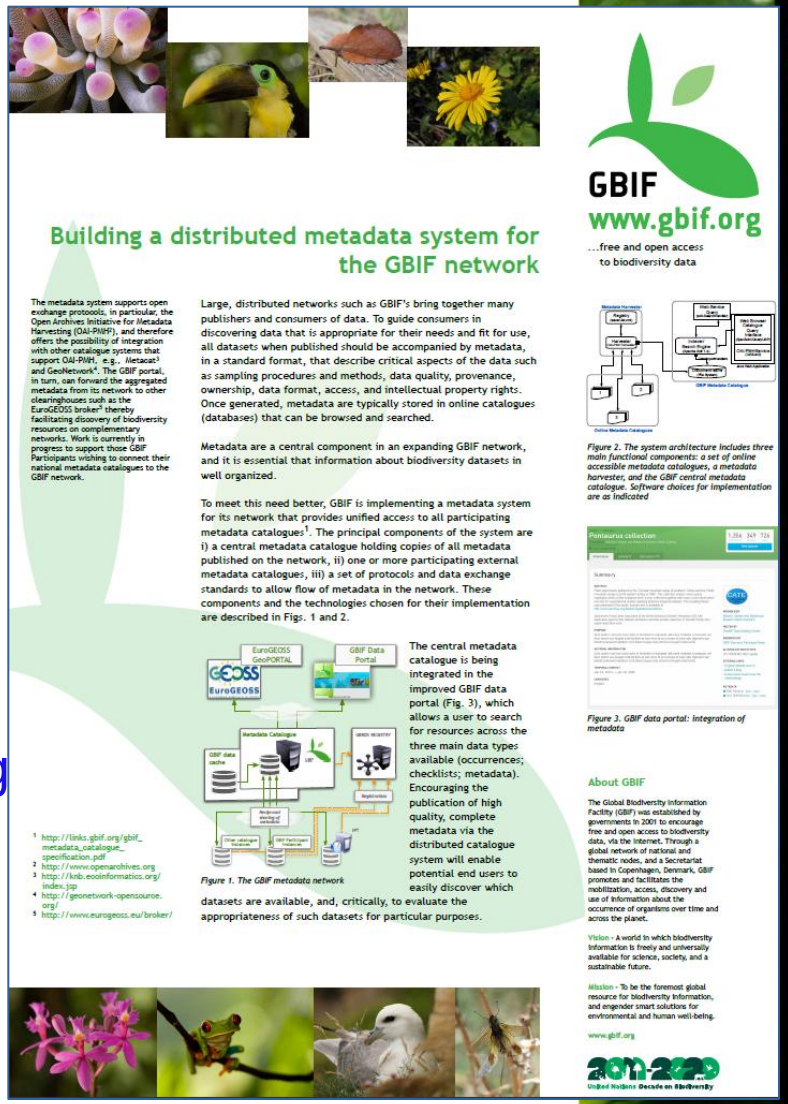
Metadata

Essential for discovery and access to data.

The GBIF metadata profile is based on Ecological Metadata Language.

OAI-PMH for interoperability across distributed metadata catalogues http://metadata.gbif.org

Output of ISO19139 to share with EuroGEOSS broker



Building a distributed metadata system for the GBIF network

The metadata system supports open exchange protocols, in particular, the Open Archives Initiative for Metadata Harvesting (OAI-PMH), and therefore offers the possibility of integration with other catalogue systems that support OAI-PMH, e.g., Metacat[3] and GeoNetwork[4]. The GBIF portal, in turn, can forward the aggregated metadata from its network to other clearinghouses such as the EuroGEOSS broker[5] thereby facilitating discovery of biodiversity resources on complementary networks. Work is currently in progress to support those GBIF Participants wishing to connect their national metadata catalogues to the GBIF network.

Large, distributed networks such as GBIF's bring together many publishers and consumers of data. To guide consumers in discovering data that is appropriate for their needs and fit for use, all datasets when published should be accompanied by metadata, in a standard format, that describe critical aspects of the data such as sampling procedures and methods, data quality, provenance, ownership, data format, access, and intellectual property rights. Once generated, metadata are typically stored in online catalogues (databases) that can be browsed and searched.

Metadata are a central component in an expanding GBIF network, and it is essential that information about biodiversity datasets in well organized.

To meet this need better, GBIF is implementing a metadata system for its network that provides unified access to all participating metadata catalogues[1]. The principal components of the system are i) a central metadata catalogue holding copies of all metadata published on the network, ii) one or more participating external metadata catalogues, iii) a set of protocols and data exchange standards to allow flow of metadata in the network. These components and the technologies chosen for their implementation are described in Figs. 1 and 2.

Figure 1. The GBIF metadata network

The central metadata catalogue is being integrated in the improved GBIF data portal (Fig. 3), which allows a user to search for resources across the three main data types available (occurrences; checklists; metadata). Encouraging the publication of high quality, complete metadata via the distributed catalogue system will enable potential end users to easily discover which datasets are available, and, critically, to evaluate the appropriateness of such datasets for particular purposes.

[1] http://links.gbif.org/gbif_metadata_catalogue_specification.pdf
[2] http://www.openarchives.org
[3] http://knb.ecoinformatics.org/index.jsp
[4] http://geonetwork-opensource.org/
[5] http://www.eurogeoss.eu/broker/

GBIF
www.gbif.org
...free and open access to biodiversity data

Figure 2. The system architecture includes three main functional components: a set of online accessible metadata catalogues, a metadata harvester, and the GBIF central metadata catalogue. Software choices for implementation are as indicated

Figure 3. GBIF data portal: integration of metadata

About GBIF

The Global Biodiversity Information Facility (GBIF) was established by governments in 2001 to encourage free and open access to biodiversity data, via the Internet. Through a global network of national and thematic nodes, and a Secretariat based in Copenhagen, Denmark, GBIF promotes and facilitates the mobilization, access, discovery and use of information about the occurrence of organisms over time and across the planet.

Vision - A world in which biodiversity information is freely and universally available for science, society, and a sustainable future.

Mission - To be the foremost global resource for biodiversity information, and engender smart solutions for environmental and human well-being.

www.gbif.org

2011-2020
United Nations Decade on Biodiversity

New data types

- complement, not duplicate work
- GBIF as top-level gateway to discovery, access

http://links.gbif.org/sp2012_2016.pdf

A roadmap developed by Q1 2013
- genomic data
- ecological data

1.3. Scop

GBIF wil

stimated 90%

t is still to be discovered

ay be 10–50 million species in

t of the undiscovered biodiversity

bit a nano-biological world

s, soils, and inside other

organisms. The currency of this knowledge will not be phenotypic data, but primarily genomic biodiversity data, with identifiers linked to animals, plants, microbes, and ecosystems.

A complementary requirement is to allow the integration of related ecological biodiversity data, as knowledge of the ecosystem function of biodiversity is critical to global and national policy and decision-making. This needs an increased focus on the relationships between species and the ecosystems they inhabit, and

Genomic Level Observations

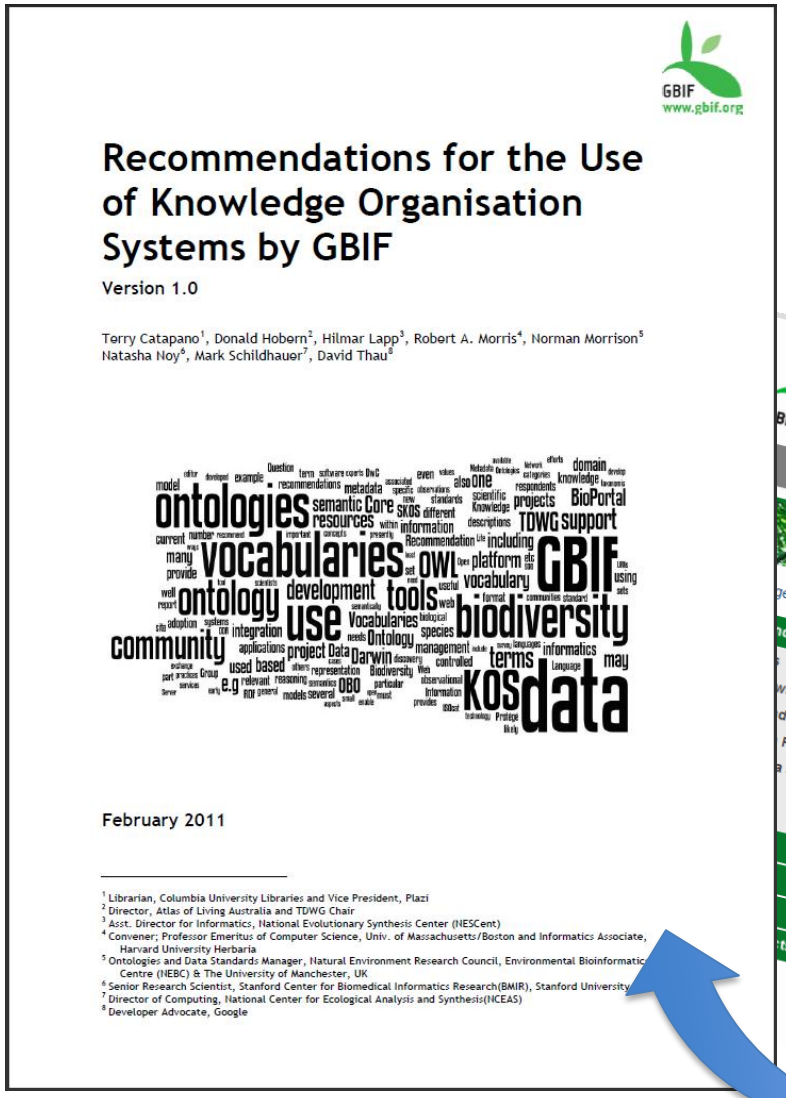
RfP for position paper on genomic level observations (Q1 2011)

Genomic Standards Consortium (GSC)
http://gensc.org

Joint activities:

- Participate in GSC biodiversity work group
- Participate in GSC conferences (Cambridge, Bremen)
- Aligning Darwin Core and GSC MIxS
- Identifying use cases,
  e.g., Microbial Earth Catalog, UNITE
- GBIF convened workshop in Q1 2012

GBIF position paper on KOS



GBIF
www.gbif.org

Recommendations for the Use of Knowledge Organisation Systems by GBIF

Version 1.0

Terry Catapano[1], Donald Hobern[2], Hilmar Lapp[3], Robert A. Morris[4], Norman Morrison[5]
Natasha Noy[6], Mark Schildhauer[7], David Thau[8]

February 2011

[1] Librarian, Columbia University Libraries and Vice President, Plazi
[2] Director, Atlas of Living Australia and TDWG Chair
[3] Asst. Director for Informatics, National Evolutionary Synthesis Center (NESCent)
[4] Convener; Professor Emeritus of Computer Science, Univ. of Massachusetts/Boston and Informatics Associate, Harvard University Herbaria
[5] Ontologies and Data Standards Manager, Natural Environment Research Council, Environmental Bioinformatics Centre (NEBC) & The University of Manchester, UK
[6] Senior Research Scientist, Stanford Center for Biomedical Informatics Research(BMIR), Stanford University
[7] Director of Computing, National Center for Ecological Analysis and Synthesis(NCEAS)
[8] Developer Advocate, Google

Robert A Morris
Terry Catapano
Donald Hobern
Hilmar Lapp
Norman Morrison
Natasha Noy
Mark Schildhauer
David Thau

GBIF
www.gbif.org

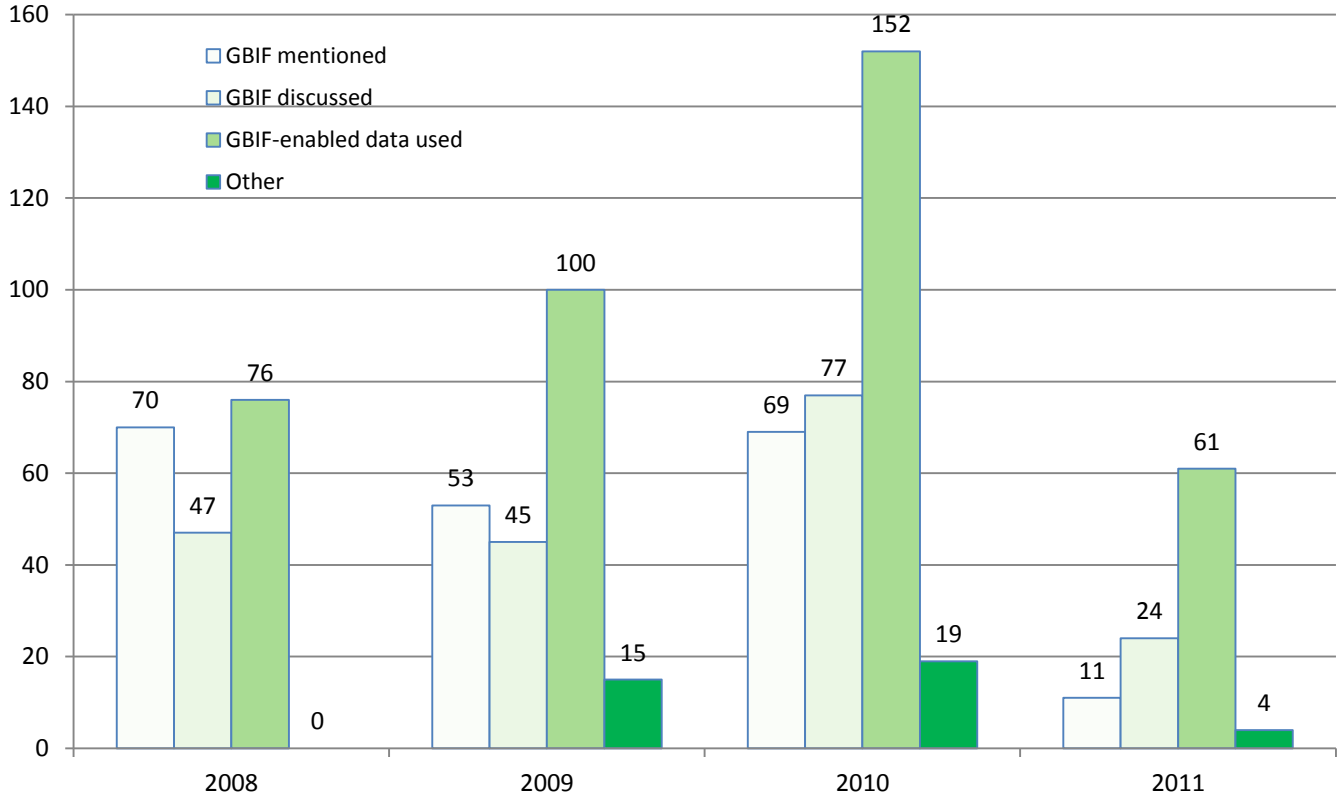
United Nations Decade on Biodiversity

Advancing scientific research – the use of GBIF mediated data

GBIF
www.gbif.org

2011-2020
United Nations Decade on Biodiversity

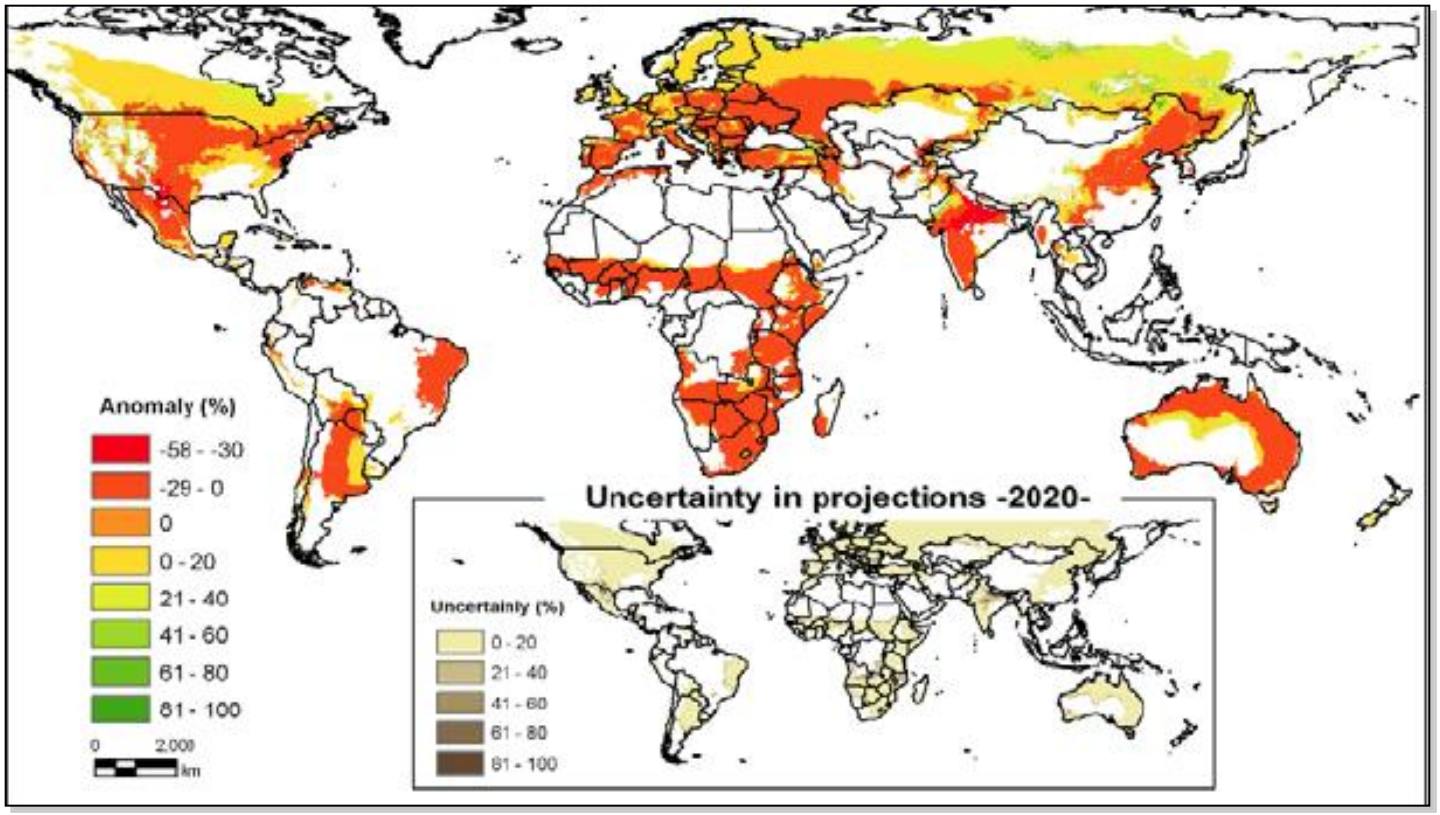
GBIF: Scientific publications



	2008	2009	2010	2011	Total
Used	76	100	152	61	**389**
Discussed	47	45	77	24	**193**
Mentioned	70	53	69	11	**203**
Other	0	15	19	4	**38**
Total	**193**	**213**	**317**	**100**	**823**

Last updated: 2011-09-22

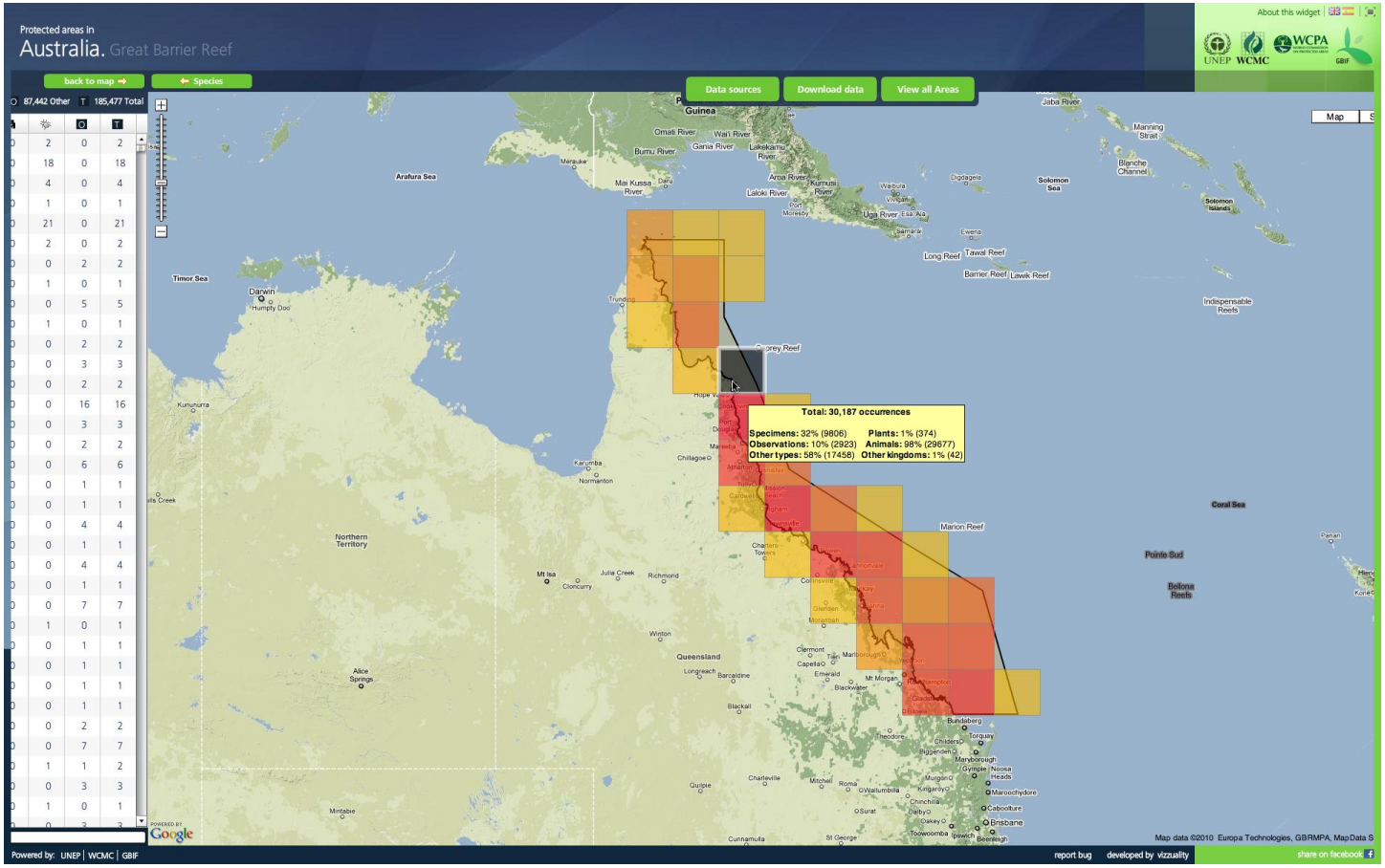
Using biodiversity data: Ecological Niche Modeling



Change in suitability for cultivating common bean across the world, from present to 2020, showing a global loss in suitability, especially in Africa.

GBIF
www.gbif.org

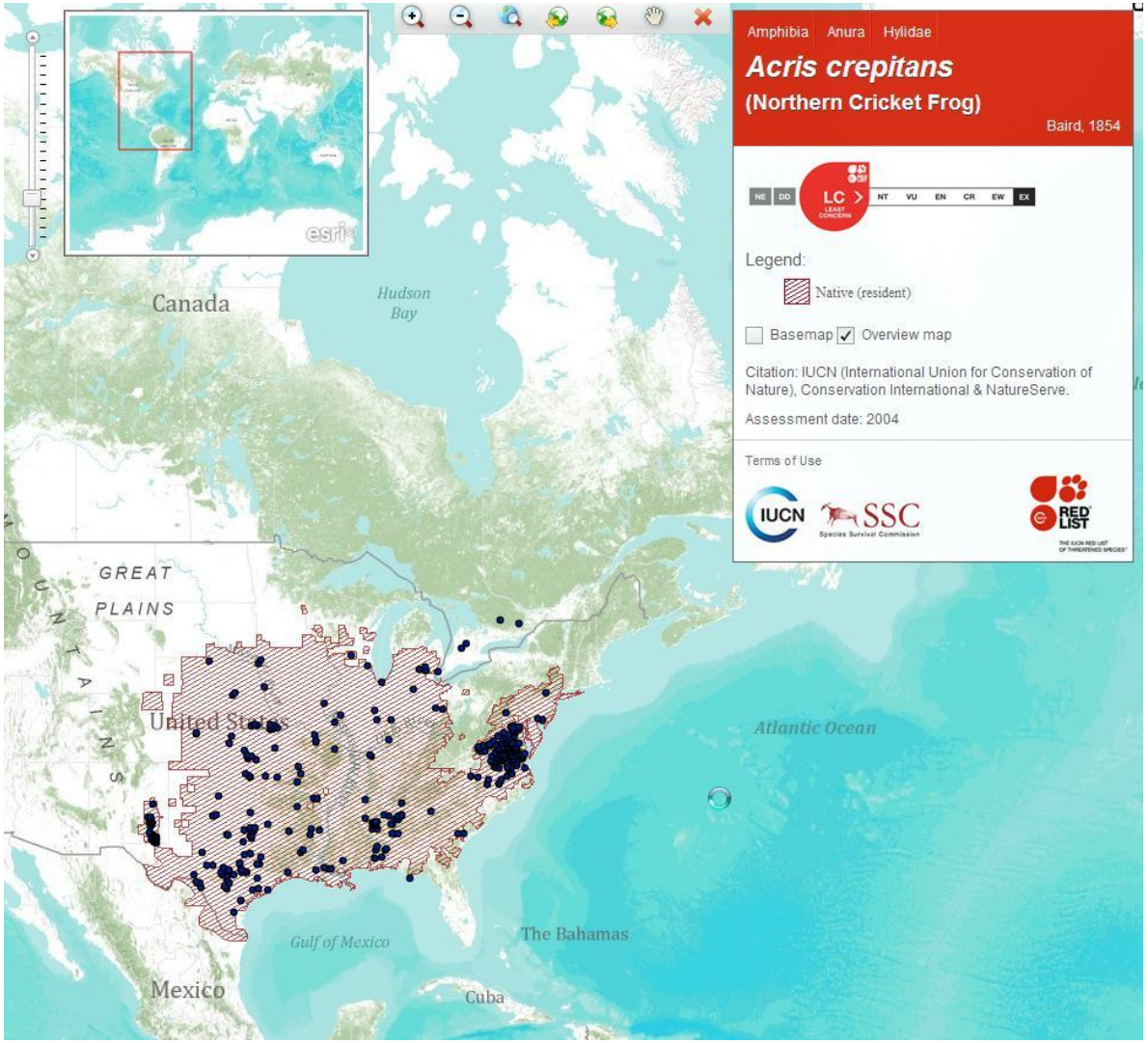
Integration: UNEP WCMC World Database on Protected Areas



GBIF
www.gbif.org

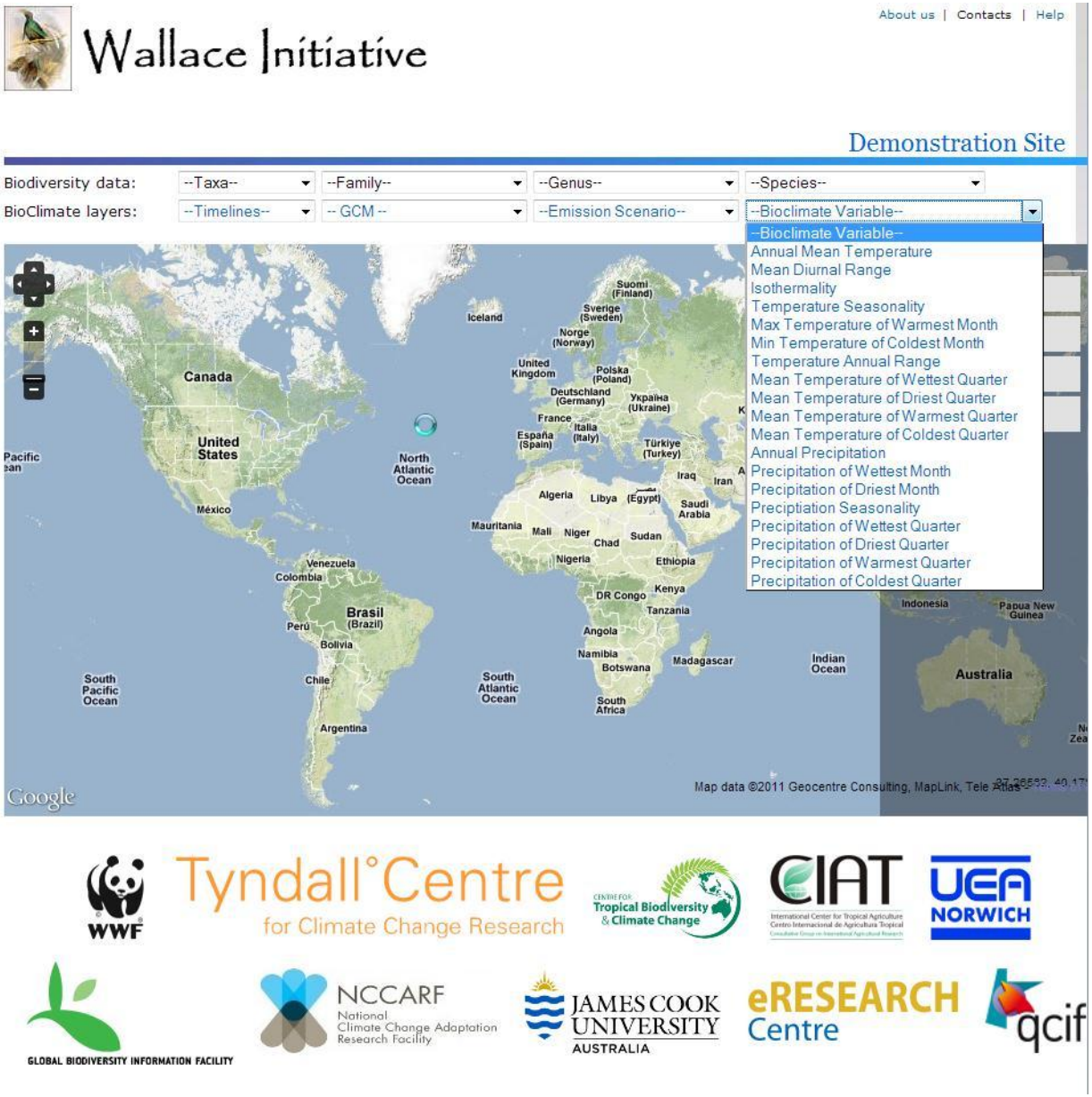
2011-2020
United Nations Decade on Biodiversity

Integration: GBIF & IUCN Red List



GBIF
www.gbif.org

2011-2020
United Nations Decade on Biodiversity



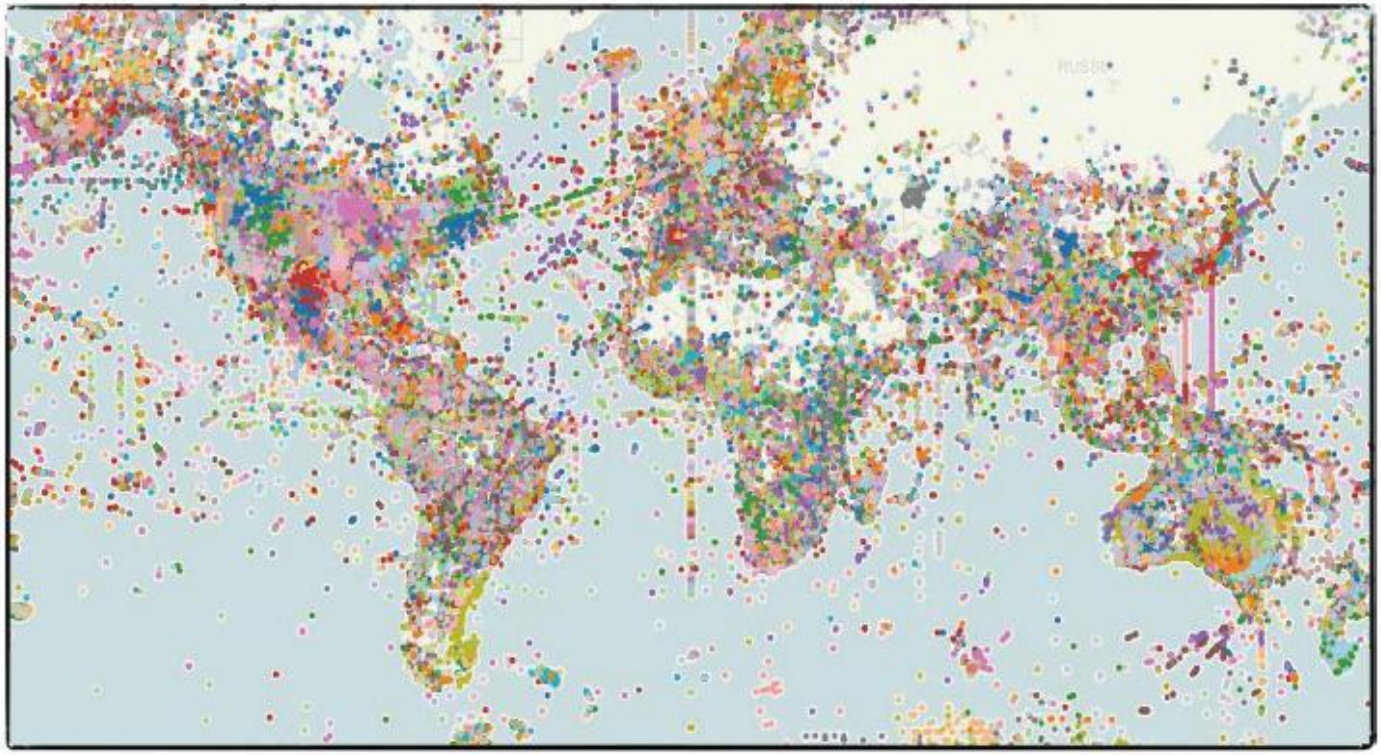
WWF
Tyndall Centre for Climate Change Research
Centre for Tropical Biodiversity & Climate Change
CIAT International Center for Tropical Agriculture
UEA Norwich
GLOBAL BIODIVERSITY INFORMATION FACILITY
NCCARF National Climate Change Adaptation Research Facility
James Cook University Australia
eRESEARCH Centre
qcif

https://wwftest.hpc.jcu.edu.au/wallace/

GBIF www.gbif.org

2011-2020 United Nations Decade on Biodiversity

CBD Access & Benefit Sharing (Nagoya) Protocol



The ABS Patent Index (ABSPAT)

Paul Oldham, Stephen Hall & Oscar Forero

ESRC Centre for Economic and Social Aspects of Genomics (Cesagen) & UNU

Image: GBIF sample distribution data for species appearing in the USPTO Patent Collection

GBIF
www.gbif.org

2011-2020
United Nations Decade on Biodiversity

ICLEI Cities Biodiversity Centre

BiodiverCities Programme

The goal of the BiodiverCities Programme is to guide, support, capacitate and motivate local governments and their partners to integrate biodiversity and ecosystem-based planning into all aspects of policy, decision making and implementation activities to result in enhanced biodiversity conservation and more sustainable local economic development.

Acknowledgement of accountability and responsibility for the health and well-being of communities and recognition of biodiversity and essential ecosystem services as the foundation of our existence are core components of the goal.

ICLEI Local Governments for Sustainability

BIODIVERCITIES TECHNICAL REFERENCE GROUP

SERVICES

- Advocacy
- Technical Support
- Profiling
- Policy Consultation

Tools & Resources

- LAB Guidebook
- TEEB Report & Manual
- Durban Commitment
- Guidelines & Case Studies
- LBSAPs
- and more... www.iclei.org/biodiversity

ENTRY POINTS FOR CITIES

- BiodiverCities
- LAB Pioneer
- LAB Pioneer Biodiversity & Climate Change
- LAB Pioneer Biodiversity & CEPA
- Cities in Biodiversity Hotspots
- URBIS
- & more to come

BiodiverCities Advisory Committee: High-level coordination group
On invitation: outstanding cities and selected organisations.

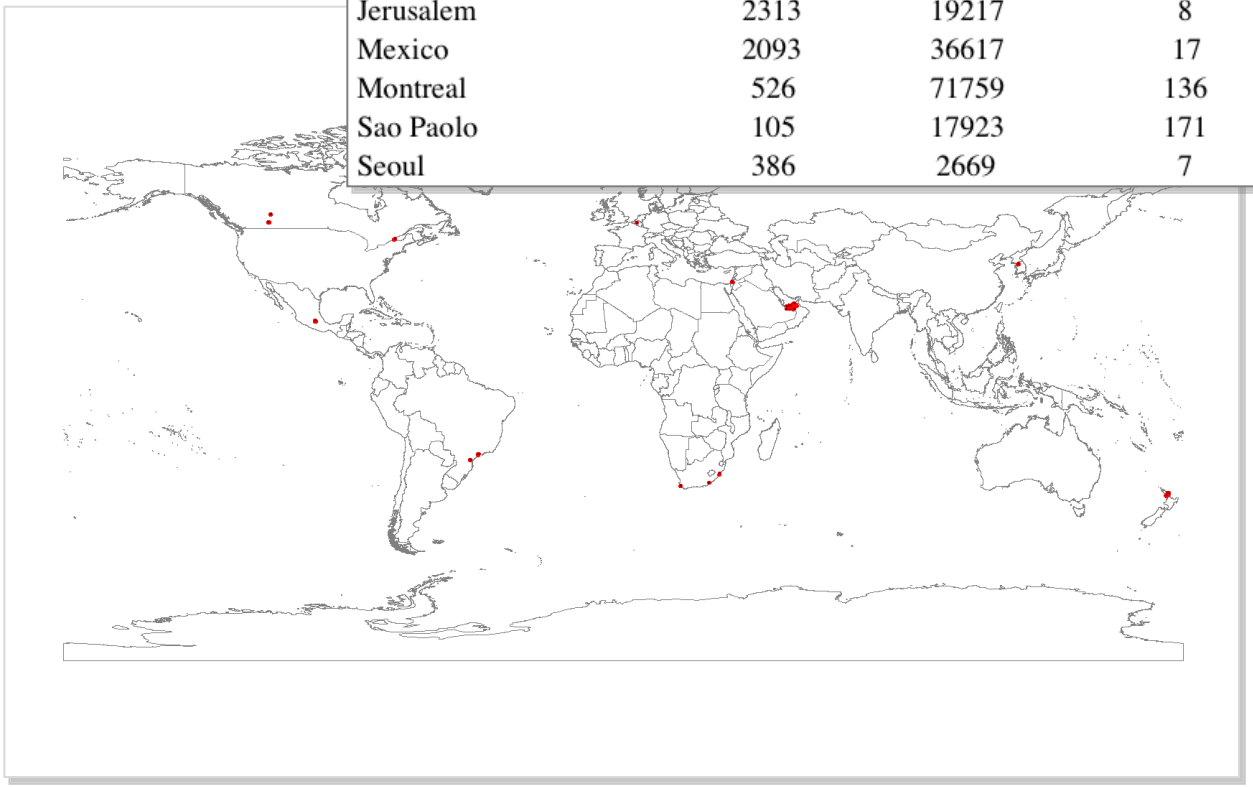
Convention on Biological Diversity | IUCN

UNEP | UN-HABITAT | National Parks – Let's Make Singapore Our Garden | GBIF | United Nations University – UNU-IAS Institute of Advanced Studies | Stockholm Resilience Centre | environmental affairs – Department: Environmental Affairs, Republic of South Africa

GLOBAL PARTNERSHIP

Accessing data from the GBIF network
(using polygons from www.gdams.org)

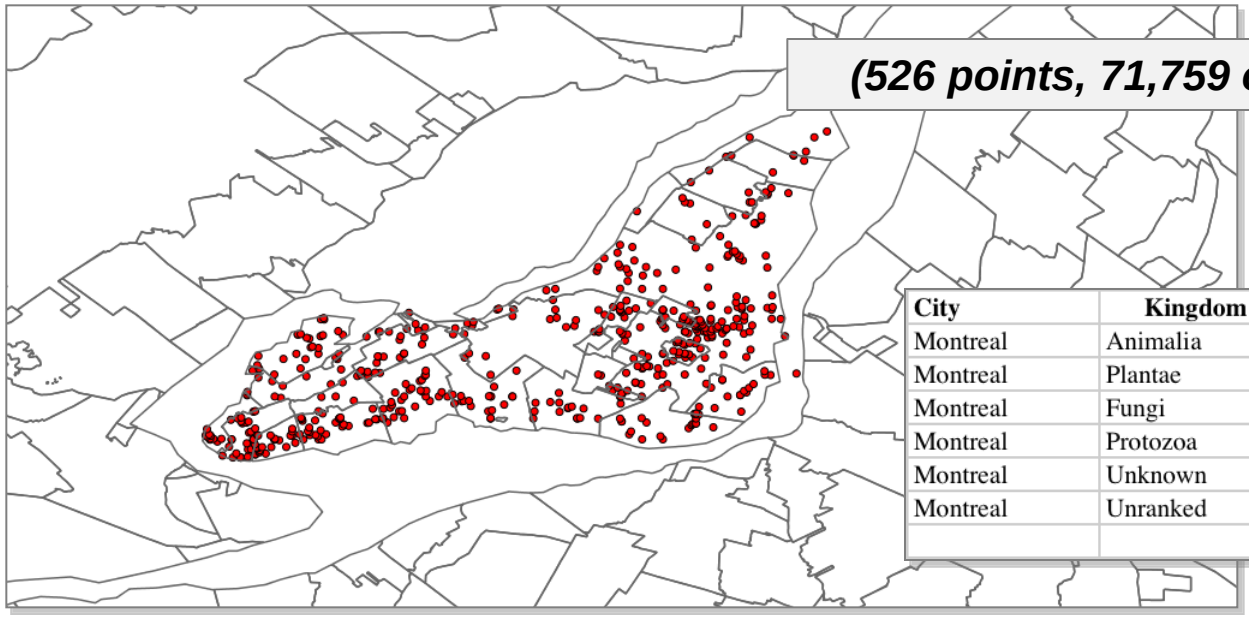
City	Points	Occurrences	Ratio
Abu Dhabi	74	502	7
Auckland-Waitakere	2166	17124	8
Brussels	148	6786	46
Calgary	542	75783	140
Cape Town	43	25614	596
Curitiba	27	877	32
Durban	118	22925	194
Edmonton	444	36525	82
Jerusalem	2313	19217	8
Mexico	2093	36617	17
Montreal	526	71759	136
Sao Paolo	105	17923	171
Seoul	386	2669	7



GBIF
www.gbif.org

2011-2020
United Nations Decade on Biodiversity

Montreal

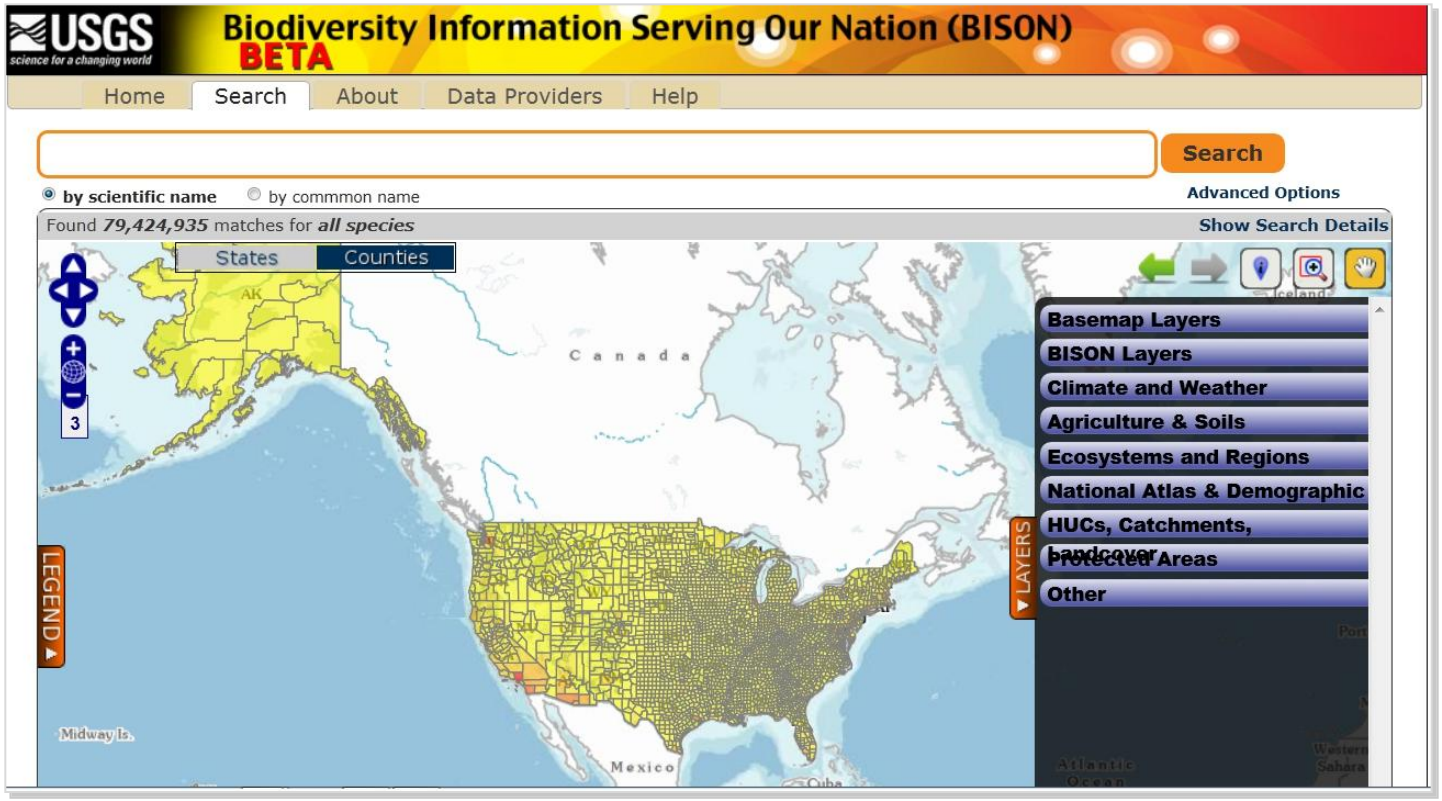


(526 points, 71,759 occurrences)

City	Kingdom	Occurrences	
Montreal	Animalia	71475	99.6%
Montreal	Plantae	245	0.3%
Montreal	Fungi	13	0.0%
Montreal	Protozoa	1	0.0%
Montreal	Unknown	5	0.0%
Montreal	Unranked	20	0.0%
		71759	

City	Kingdom	Phylum	Class	Order	Family	*Genus*	*Species*	Occurrenc
Montreal	Animalia	Chordata	Aves	Passeriformes	Paridae	Poecile	atricapillus	5093
Montreal	Animalia	Chordata	Aves	Passeriformes	Passeridae	Passer	domesticus	4577
Montreal	Animalia	Chordata	Aves	Passeriformes	Fringillidae	Cardinalis	cardinalis	3810
Montreal	Animalia	Chordata	Aves	Passeriformes	Sturnidae	Sturnus	vulgaris	3498
Montreal	Animalia	Chordata	Aves	Passeriformes	Fringillidae	Carduelis	tristis	3485
Montreal	Animalia	Chordata	Aves	Columbiforme	Columbidae	Zenaida	macroura	3265
Montreal	Animalia	Chordata	Aves	Piciformes	Picidae	Picoides	pubescens	3072
Montreal	Animalia	Chordata	Aves	Passeriformes	Fringillidae	Carpodacus	mexicanus	2654
Montreal	Animalia	Chordata	Aves	Passeriformes	Corvidae	Corvus	brachyrhyncho	2637
Montreal	Animalia	Chordata	Aves	Columbiforme	Columbidae	Columba	livia	2310

Building national data portals



- Increased geospatial granularity (national to county)
- Nationally-relevant thematic layers

GBIF
www.gbif.org

GBIF and GEO BON

GBIF
www.gbif.org

2011-2020
United Nations Decade on Biodiversity

GBIF and GEO BON

Co-lead of WG8 on Data Integration & Interoperability

Key outputs:

1. Detailed Implementation Plan
2. Principles of the GEO BON Information Architecture

GBIF
www.gbif.org

2011-2020
United Nations Decade on Biodiversity

GEO BON Detailed Implementation Plan

Concepts to be implemented

- networks and their information resources
- data types and data content
- existing global networks
- national and regional networks
- discovery services and registries
- interoperability and information management services
- ontologies, thesauri, dictionaries, semantic mediation
- organism names and habitat classifications
- workflow of services and integration of applications
- portals, search engines, querying and harvesting
- open access issues

GBIF
www.gbif.org

United Nations Decade on Biodiversity

GEO BON Detailed Implementation Plan

Activities

- Establish a working group and coordinating unit for technical implementation
- Review existing data provider networks and establish partnerships
- Review of data processing needs of other WGs
- Design the information architecture of GEO BON
- Build the components such as portal, registry, ontologies
- Register data and services and provide helpdesk
- Outreach and capacity building

GBIF
www.gbif.org

United Nations Decade on Biodiversity

Principles of the GEO BON Information Architecture (1)

(companion document to the Detailed Implementation Plan)

- Documents the "*diversity of biodiversity networks*" and their chief characteristics;
- Highlights how GEO BON can *leverage* the work of *existing networks and initiatives*;
- Proposes approach to *informatics design* based on a Service Oriented Architecture as described in the *GEOSS Common Infrastructure*.

.

GBIF
www.gbif.org

United Nations Decade on Biodiversity

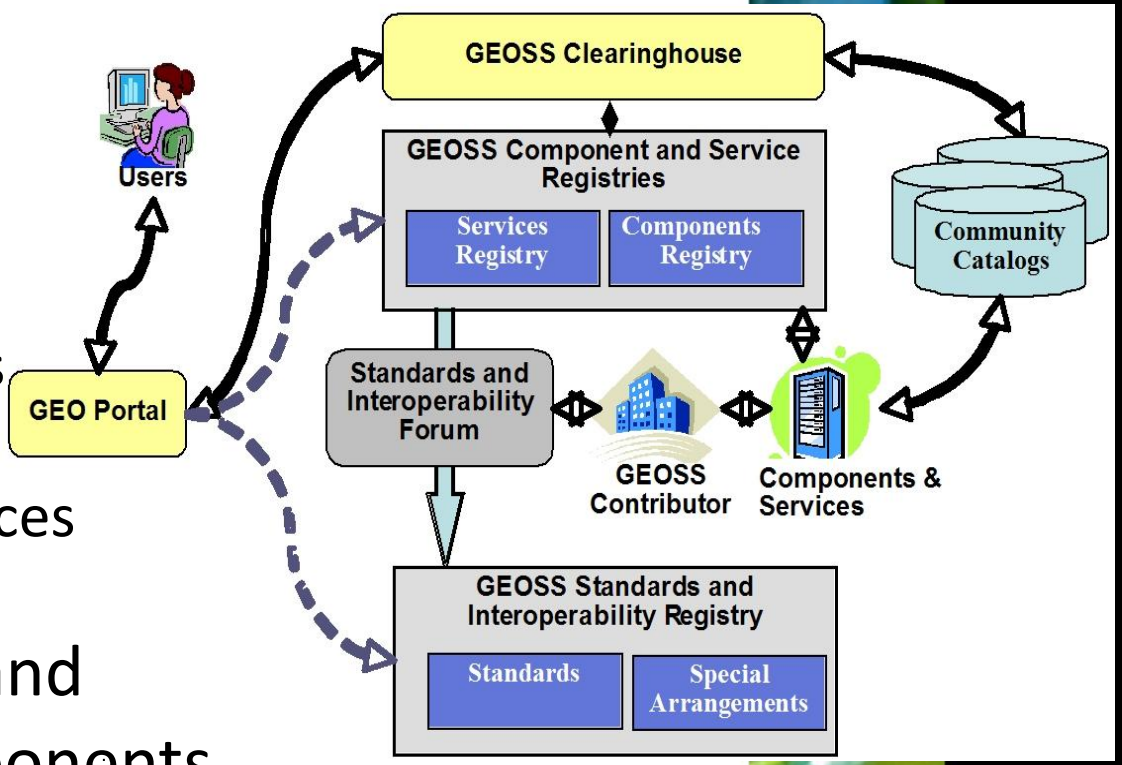
GEOSS Common Infrastructure

A Service Oriented Architecture featuring:

- loose coupling of services

- consumers discover services

Metadata catalogues and registries are key components



Interactions of GEOSS Registries, Portal and Clearinghouse

Extracted from: *GEOSS Core Architecture Implementation Report (http://portal.opengeospatial.org/files/?artifact_id=24315)*

Principles of the GEO BON Information Architecture (2)

Covers:

- 41 existing global, regional and national networks, discovery services and registries (GBIF, ILTER, KNB, NASA GCMD, NBII);
- 20 standards (metadata, data, transfer protocols);
- Ontologies, vocabularies for semantic mediation;
- Biological names and habitat classifications;
- Workflow of services and integration of applications (climate change scenario example using GBIF data);
- Portals, search engines, querying and harvesting including GBIF Data Portal, LTER/ILTER, NBII, KNB and NASA GCMD facilities;
- Open Access Issues and GEOSS Data Sharing Principles.

GBIF
www.gbif.org

United Nations Decade on Biodiversity

GEOSS and data sharing

"The societal benefits of Earth observations cannot be achieved without data sharing“
The 10-Year Implementation Plan

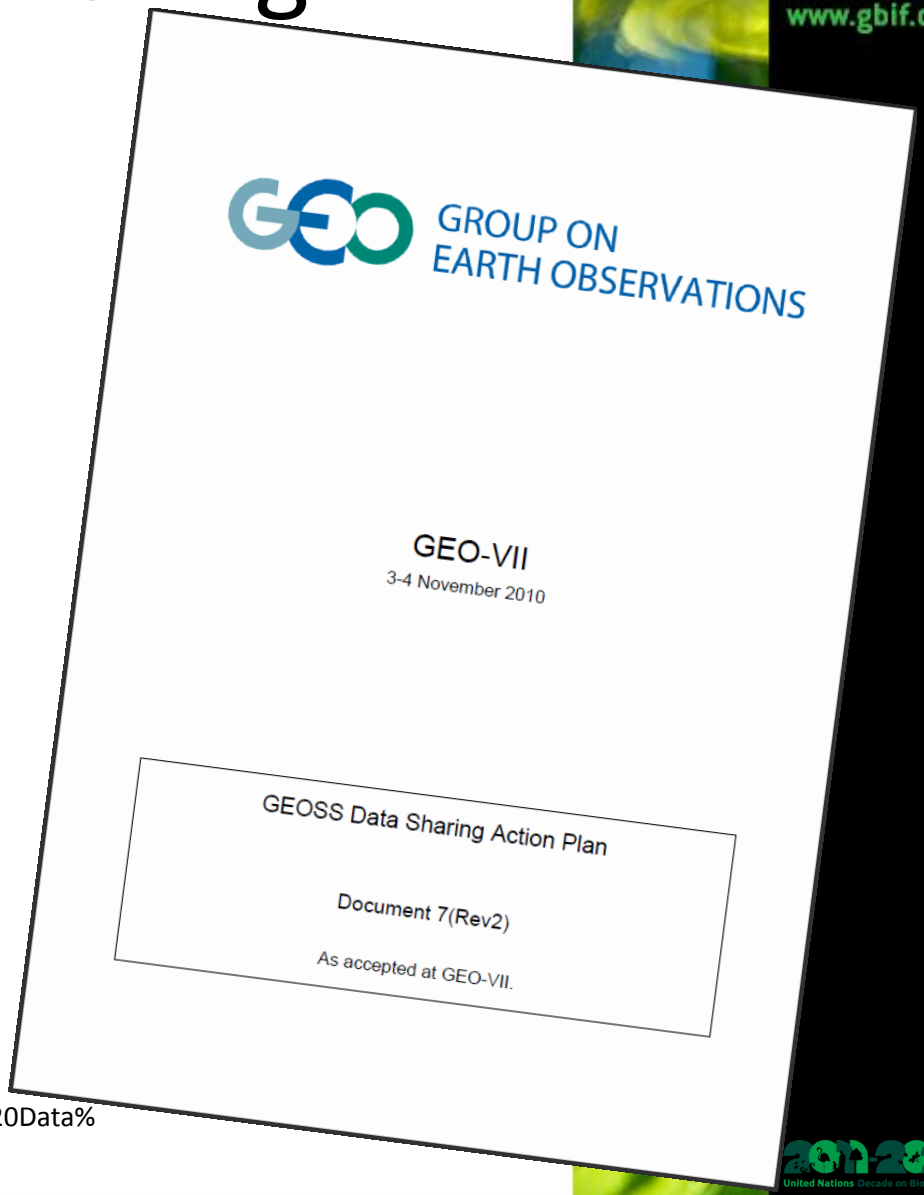
“... full and open exchange of data, metadata and products shared within GEOSS, recognizing relevant international instruments and national policies and legislation”

GBIF
www.gbif.org

United Nations Decade on Biodiversity

GEOSS and data sharing

GBIF is one of the case studies in the GEOSS Data Sharing and Action Plan promoting full and open sharing of data.



http://www.earthobservations.org/documents/geo_vii/07_GEOSS%20Data%20Sharing%20Action%20Plan%20Rev2.pdf

geossregistries.info/geosspub/

GROUP ON EARTH OBSERVATIONS

Feedback for this page

Back

GEOSS Registry Publication Portal

GEOSS Component Details

Component Basic Information

Component Id:	urn:uuid:54374b70-25e6-4071-a530-b137bdcb130b
Name:	Global Biodiversity Information Facility Data Portal
Abbreviation:	GBIF Data Portal
Description:	GBIF is an international organisation that is working to make the world's biodiversity data freely accessible. The GBIF data portal is a service that provides access to millions of data records that are being shared via the GBIF
GEO Member or Participating Organisation :	
Responsible Organisation:	
URL to Component Information:	

Component Contact Information

Contact Name:	
Contact Email:	

Component Category

websitesDocuments

Societal Benefit Areas

- Agriculture
- Biodiversity
- Climate
- Ecosystems
- Health
- Water

GROUP ON EARTH OBSERVATIONS

Feedback for this page

Back

GEOSS Registry Publication Portal

Associated GEOSS Services for GEOSS Component: Global Biodiversity Information Facility Data Portal

Associated Services

1.	Global Biodiversity Information Facility Network Web Service	Details
2.	Global Biodiversity Information Facility Provider Web Service	Details
3.	Global Biodiversity Information Facility Occurrence Web Service	Details
4.	Global Biodiversity Information Facility Density Web Service	Details
5.	Global Biodiversity Information Facility Taxon Web Service	Details
6.	Global Biodiversity Iinformation Facility Resource Web Service	Details

Back

Last updated: Friday, November 4, 2011

GBIF www.gbif.org

EuroGEOSS Broker

GBIF provides services for the EuroGEOSS broker

	CSW	WMS	WFS	Others
Forest	-	4 services (38 datasets)	-	-
Biodiversity	-	-	1 service	1 GBIF-service
Drought	2 services (102 datasets)	6 services (161 datasets)	3 services (40 datasets)	1 WFS-G
Generic	2 services	-	-	-

CSW - « Catalogue Service for the Web »
Service used to request metadata catalogues of datasets and services.

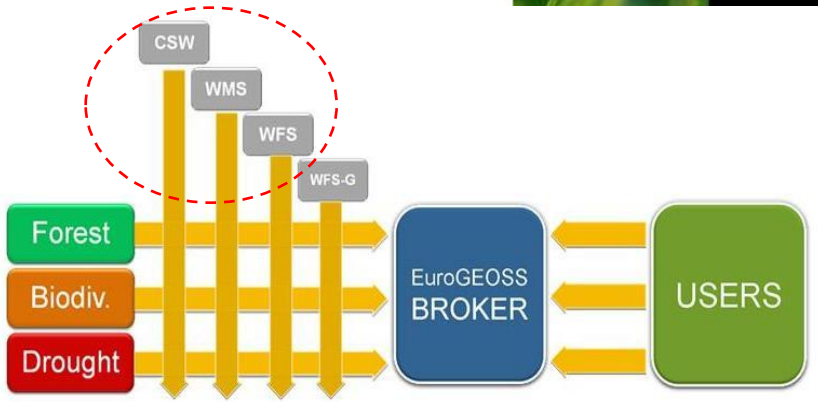
WMS - « Web Map Service »
Service used to download geospatial information in a raster format. WMS are mainly view services.

WFS - « Web Feature Service »
Service delivering raw geospatial data (under GML). WFS are mainly downloading services.

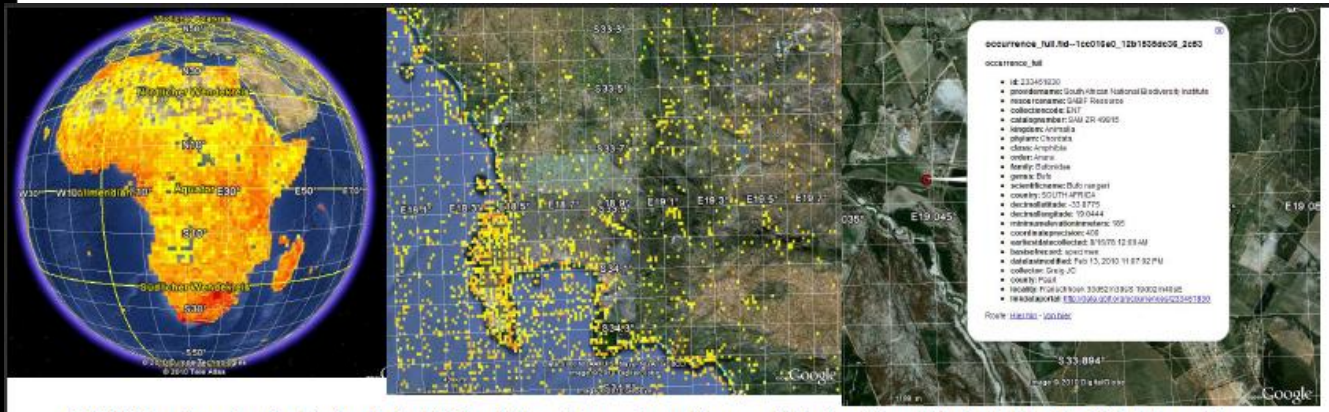
WFS-G
WFS used to deliver gazetteer service (producing Bounding Box from toponyms).

GBIF
Specific query interface connected to the GBIF (Global Biodiversity Information Facility) metadata catalogue.

- *Web Map Service*
- *Web Feature Service*
- *~~CSW~~ -> OAI-PMH*



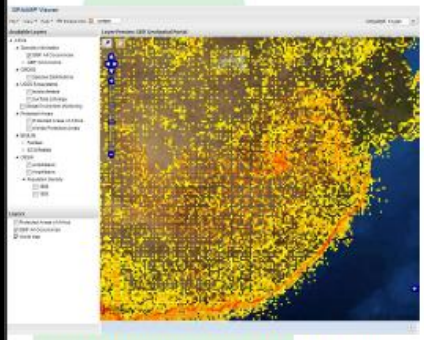
Source: http://www.eurogeoss.eu/broker/default.aspx



GBIF OGC services visualised in Google Earth. The rightmost image shows the record details with a clickable link to the GBIF data portal.
GBIF WMS African Biota - http://ogc.gbif.org/wms?service=WMS&version=1.1.1&request=GetCapabilities
GBIF WFS African Biota - http://ogc.gbif.org/wfs?service=WFS&version=1.1.0&request=GetCapabilities

OGC Web Services for GBIF-Mediated Occurrence Data

GBIF
www.gbif.org
...free and open access to biodiversity data



The Open Geospatial Consortium (OGC) is a non-profit, international, voluntary consensus standards

One of the goals of the EU-funded EuroGEOSS[1] project is to develop the infrastructure for a "biodiversity operational capacity" as a European contribution to the Group on Earth Observations Biodiversity Observation Network (GEO BON)[2]. One outcome of EuroGEOSS will be the Digital Observatory for Protected Areas (DOPA[3]), an information system for assessing the state and pressure of protected areas in order "to support proper prioritisation for decision making and fund allocation processes". As a contribution to the DOPA, GBIF developed a number of OGC[4] web services facilitating geospatial access to GBIF-mediated African biodiversity data.

GBIF Contacts
GIS Developer, Jörg Holetschek,
j.holetschek@bgbm.org

Information Systems Architect,
Tim Robertson, trobertson@gbif.org

Senior Programme Officer, Inventory, Discovery, Access (IDA),
Éamon Ó Tuama, eotuama@gbif.org

dopa.jrc.ec.europa.eu

DOPA

Digital Observatory for Protected Areas

Home | Overview | Projects | Web Services | Publications & Resources | News & Events | Who Are We?

Search

USER LOGIN

Username: *

Password: *

Log in

Request new password

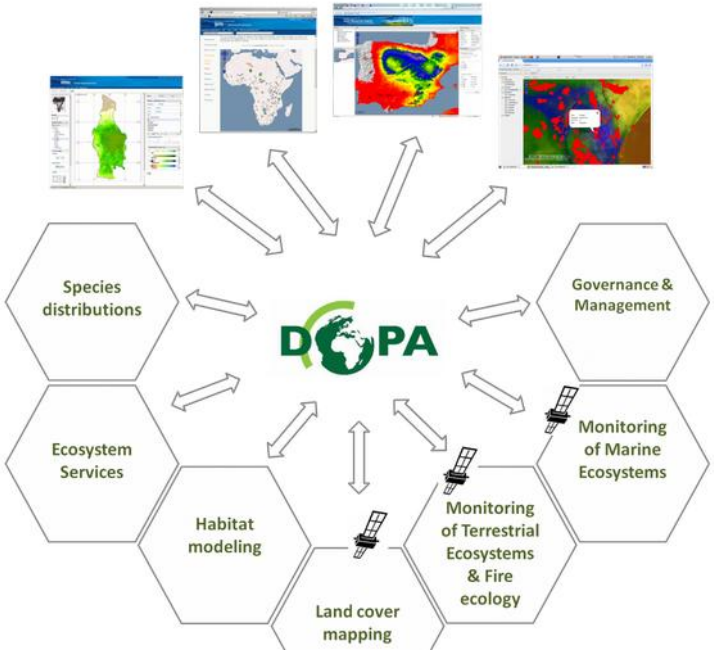
RECENT CHANGES

Web Services & Clients
(9 weeks 6 days ago)
Who are we?
(11 weeks 1 hour ago)
Projects
(13 weeks 3 days ago)
Publications & Resources
(15 weeks 6 days ago)
The Digital Observatory for Protected Areas
(16 weeks 8 hours ago)
An introduction to the DOPA
(18 weeks 4 days ago)
News & Events
(18 weeks 4 days ago)

The Digital Observatory for Protected Areas

The Digital Observatory for Protected Areas (DOPA) has been created as a component of the GEO-BON observation network by the Joint Research Centre in collaboration with other international organizations including the Global Biodiversity Information Facility (GBIF), the UNEP-World Conservation Monitoring Centre (WCMC), Birdlife International and the Royal Society for the Protection of Birds (RSPB). DOPA is conceived as a set of distributed databases combined with open, interoperable web services (Figure 1) to provide a large variety of end-users including park managers, decision-makers and researchers with means to assess, monitor and forecast the state and pressure of protected areas at the global scale.

DOPA is also a contribution to the Group on Earth Observations Biodiversity Observation Network (GEO BON), the biodiversity arm of the Global Earth Observation System of System of Systems (GEOSS).



ADDTOANY

Share / Save

GBIF
www.gbif.org

2011-2020
United Nations Decade on Biodiversity

GBIF
www.gbif.org

2011-2020
United Nations Decade on Biodiversity

Key messages

- Highly complex, non-trivial challenges – people, data, infrastructure;
- 10 years experience to date, and much already accomplished, allowing GEO BON to ‘jump start’ on GBIF;
- Based on partnerships, common purpose and ‘public good’ philosophy;
- GBIF mediated data easily deployed for GEOSS, e.g., Architecture Implementation Pilot studies, DOPA, EuroGEOSS broker.
- Watch this space!!!

GBIF
www.gbif.org

United Nations Decade on Biodiversity

GLOBAL
BIODIVERSITY
INFORMATION
FACILITY

GBIF Strategic Plan 2012-2016

Seizing the Future

WWW.GBIF.ORG

GBIF
www.gbif.org

2011-2020
United Nations Decade on Biodiversity