

GBIF and IAVD

Finn Borchsenius
Department of Bioscience
Aarhus University



You are not logged in [[Login](#)]

[GBIF](#) | [GBIF data portal](#) | [allearter.dk](#) | [Naturhistorisk Guide](#) | [Centre for Invasive Species](#) | [Danmarks Miljøportal](#) | [DK-CHM](#) | [CBD](#)



DanBIF

Danish Biodiversity Information Facility



[HOME](#) | [LOGIN](#) | [SEARCH](#) | [FORUM](#) | [NEWS](#) | [EVENTS](#)

[En](#) | [Da](#)

About DanBIF 
[Contact](#)
[Organisation](#)
[Objective and work](#)

Products 
[DK datasets in GBIF](#)
[Data catalogue](#)
[Meetings, conferences](#)
[Communication](#)
[Digitisation projects](#)

Services 
[GBIF Data Use Agreement](#)
[GBIF Data User Guide](#)
[Digitisation](#)
[Georeferencing](#)
[Our websites](#)
[DanBIF web service](#)

[Home](#)

DanBIF is the Danish national node of GBIF - Global Biodiversity Information Facility

DanBIF is also a [GBIF](#) node for the Faroe Islands and Greenland.

DanBIF is funded by the The Danish Council for Independent Research – Natural Sciences of the Ministry of Science, Technology and Innovation, and The Faculty of Science, University of Copenhagen.

The international megascience initiative GBIF aims to make all the world's information on biodiversity freely and universally available for query and analysis via the Internet. DanBIF carries out the work on making accessible the Danish, Faroese and Greenlandic resources of biodiversity data. DanBIF is also a communication forum for a Danish network of biodiversity researchers. Read more about GBIF and DanBIF [here](#).

See the overview of [data sets from Denmark](#) provided to the GBIF data portal.

DanBIF Collections Metadata Inventory - survey of collections in Denmark

Search for collections in The DanBIF [Collections Metadata Inventory](#).

You can also register your collections in the Metadata Inventory, read more [here](#).

DANBIF IN NUMBERS

Data records: 7.840.508

92% concerning Danish species

Datasets 52

Institutions: 65

DanBIF network: 190 pers.

GBIF DATA AGREEMENTS

For use of GBIF data see [GBIF Data Use Agreement](#).

For sharing data see [GBIF Data Sharing Agreement](#).

LATEST DATASETS

[Aves Tanzanian collection at the Natural History Museum of Denmark \(SNM\)](#)

[Bugbase](#) (Lepidopterological Society)
[Vascular plants in Denmark recorded under The Nationwide Monitoring and Assessment Programme for the](#)



About GBIF

The Global Biodiversity Information Facility

The OECD origin...

“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)”



The Global Biodiversity Information Facility

Vision

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

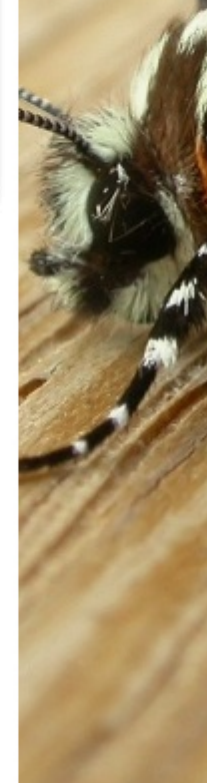
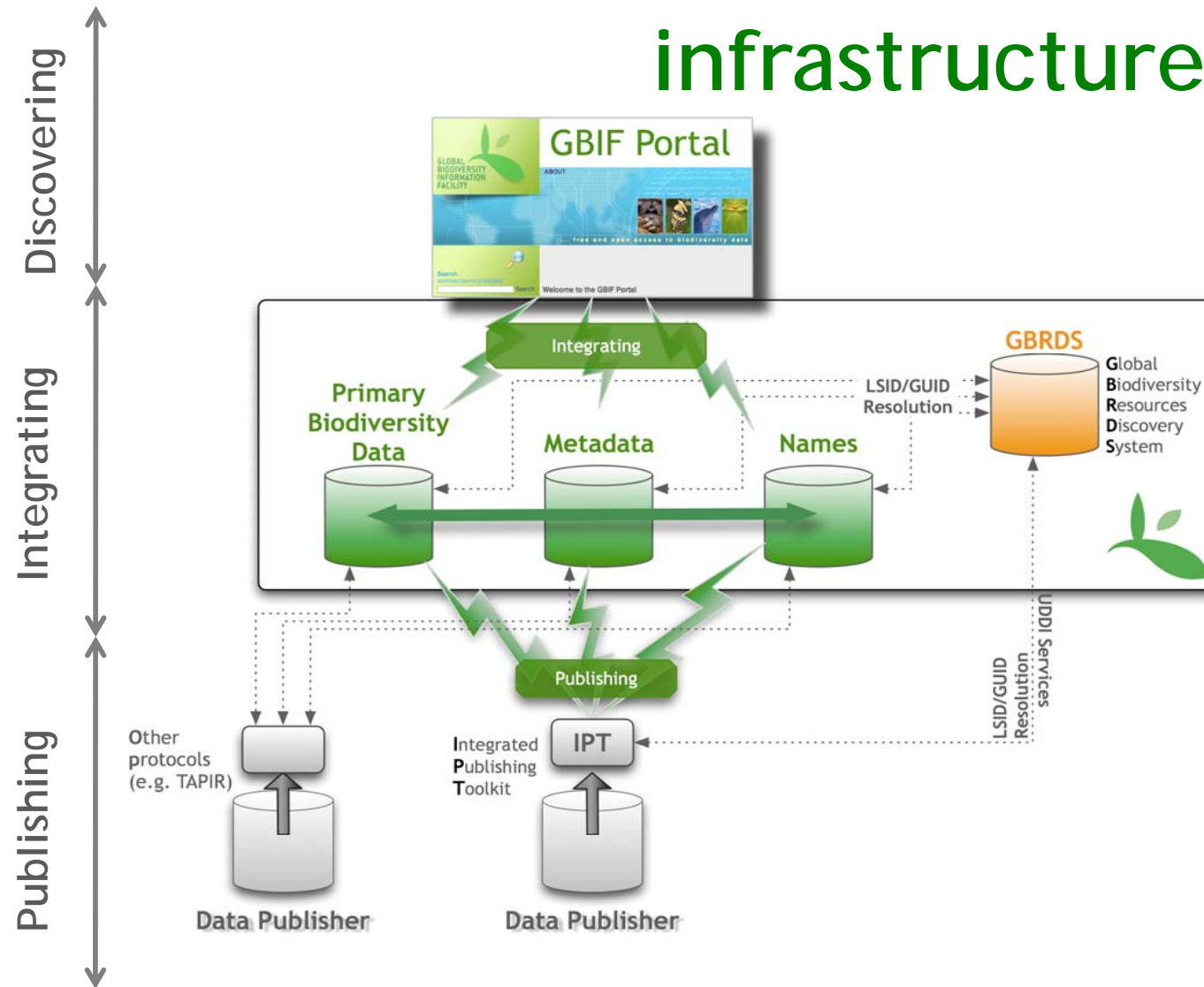


Mission

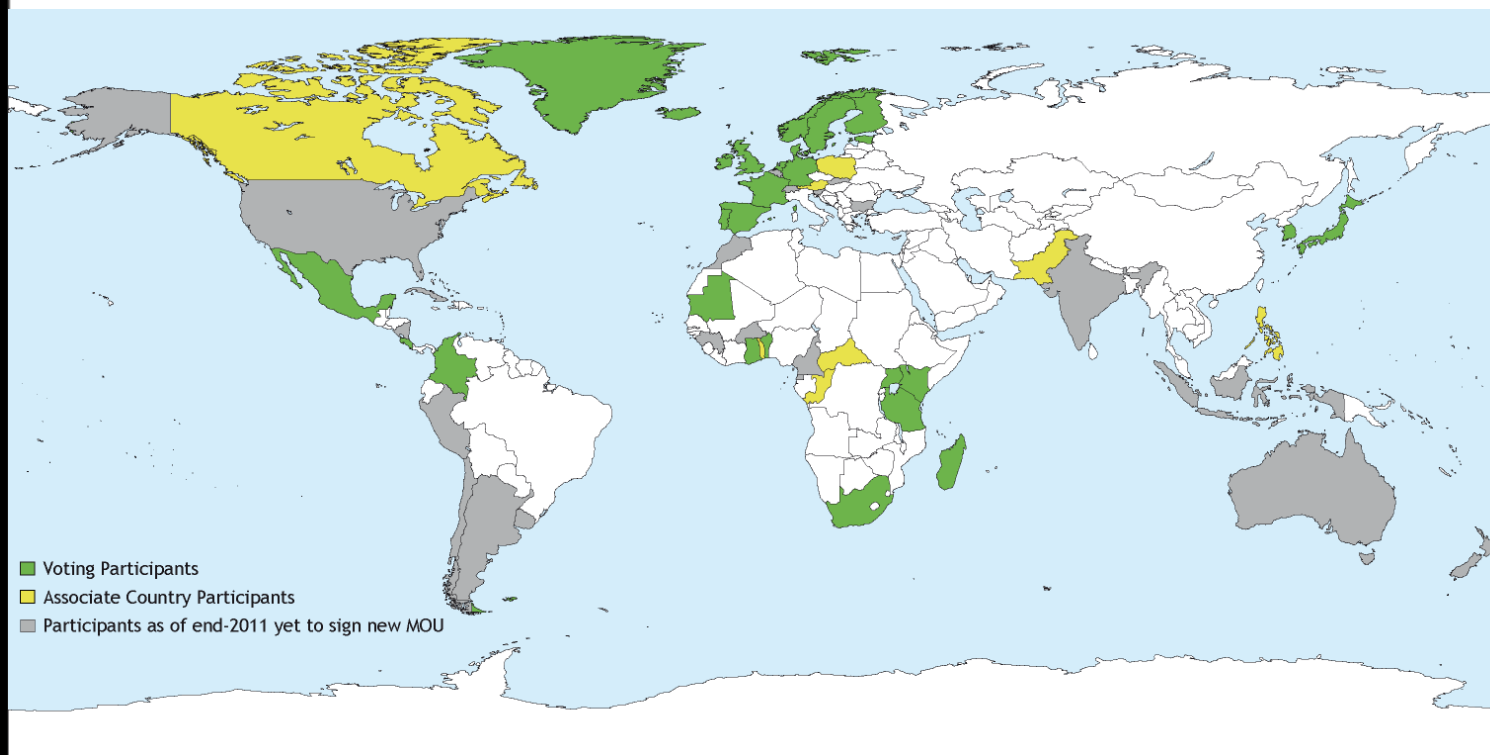
To be the foremost global resource for biodiversity information, and engender smart solutions for environmental and human well-being.



The GBIF informatics infrastructure



GBIF Participation



About the data

Primary Biodiversity Data

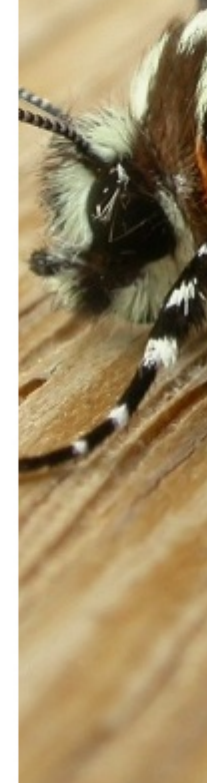
Primary Biodiversity Data is defined as: Digital text or multimedia data record detailing facts about the instance of occurrence of an organism, i.e. on the what, where, when, how and by whom of the occurrence and the recording.



Observational data



Specimen data



How to publish biodiversity data

Darwin Core

The purpose of DwC terms is to facilitate data sharing

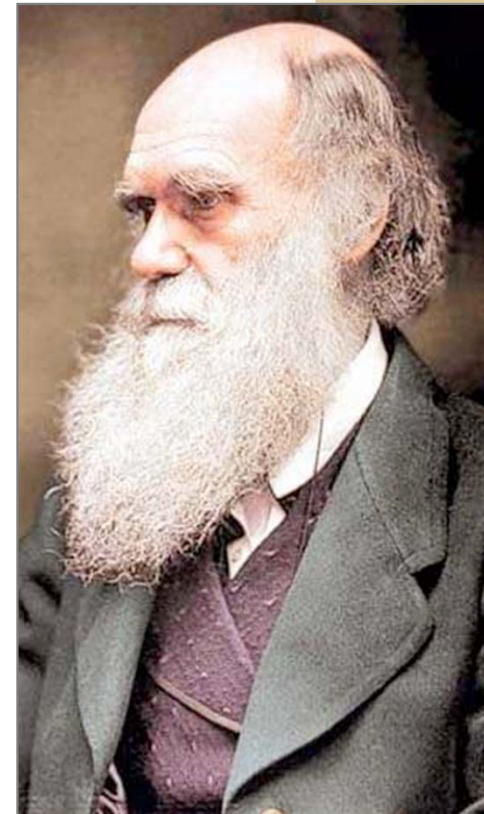
- a well-defined standard core vocabulary
- a flexible framework to maximize re-usability

The Darwin Core can be extended by adding new terms to share additional information.

Approved as TDWG standard 2009

“The Darwin Core is primarily based on taxa, their occurrence in nature as documented by observations, specimens, and samples, and related information.”

<http://rs.tdwg.org/dwc/>



How to publish biodiversity data

Darwin Core

The Darwin Core was specifically designed to facilitate the exchange of information about the geographic occurrence of organisms and the physical existence of biotic specimens in collections. Extensions to the Darwin Core provide a mechanism to share additional information, which may be discipline-specific, or beyond the current agreed scope of the Darwin Core itself.

DwC Taxon Core Record

dwc:datasetID	ID referring to dataset and its metadata	string
dwc:taxonID	LSID, GUID or even local ID.	string
dc:source	Link to human readable webpage	uri
dc:accessRights	access right details	string
dc:modified	date modified	date
dc:rights	IPR	string
dc:rightsHolder		string
dc:language	language of the source language	voc
dwc:kingdom	Kingdom taxon group name	string
dwc:phylum	Phylum taxon group name	string
dwc:class	Class taxon group name	string
dwc:order	Order taxon group name	string
dwc:family	Family taxon group name	string
dwc:genus	Genus group name	string
dwc:subgenus	SubGenus group name	string
dwc:specificEpithet	species name	string
dwc:infraspecificEpithet	infraspecies name	string
dwc:scientificName	Full unparsed scientific name	string
dwc:taxonRank	Taxon Rank Name from vocabulary	voc
dwc:scientificNameAuthorship	authorship of terminal name	string
dwc:higherTaxonID	ID as ForeignKey	string
dwc:higherTaxon	ScientificName as ForeignKey	string
dwc:acceptedTaxonID	ID as ForeignKey	string
dwc:acceptedTaxon	ScientificName as ForeignKey	string
dwc:basionymID	ID as ForeignKey	string
dwc:basionym	ScientificName as ForeignKey	string
dwc:namePublishedIn	Bibliographic citation of nomenclatural act	string
dwc:taxonAccordingTo	Bibliographic citation of usage	string
dwc:taxonomicStatus	Taxonomic status of record	voc
dwc:nomenclaturalStatus	Nomenclatural status of record	string
dwc:nomenclaturalCode	Nomenclatural Code (ICBN)	string
dwc:taxonRemarks	Notes	string

Core Record



Darwin Core Archive (DwC-A)

A standard to allow efficient sharing of

- i. Occurrence based data
- ii. Taxonomic checklists

Flexibility to support custom extensions such as

- i. Distribution information
- ii. References
- iii. Images
- iv. ... *create your own*

Includes dataset descriptive “metadata”

- i. Sampling methods, quality control
- ii. Citations
- iii. ... *etc*



DwC-A related references

Fauna Europaea checklist example

- http://ecat-dev.gbif.org/repository/protected/taxonomic/fauna_europaea.zip

Darwin Core Archive introduction

- http://www.gbif.org/orc/?doc_id=2819&l=en

How to create a Darwin Core Archive

- http://www.gbif.org/orc/?doc_id=2816&l=en

Darwin Core Archive extensions

- <http://rs.gbif.org/>

Darwin Core validator service

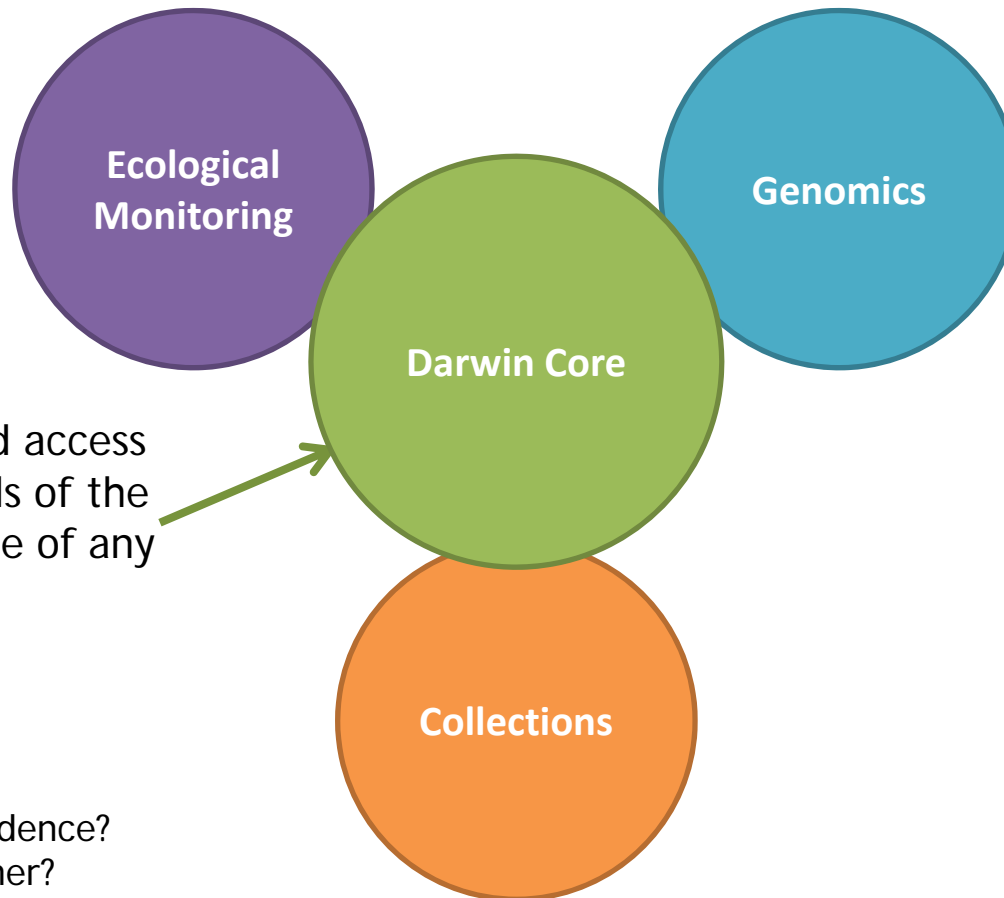
- <http://tools.gbif.org/dwca-validator/>

GBIF Metadata Profile

- http://www.gbif.org/orc/?doc_id=2820



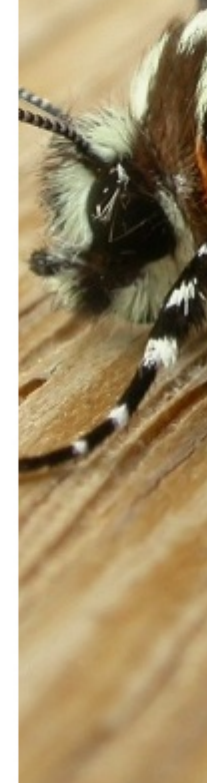
Unifying species data



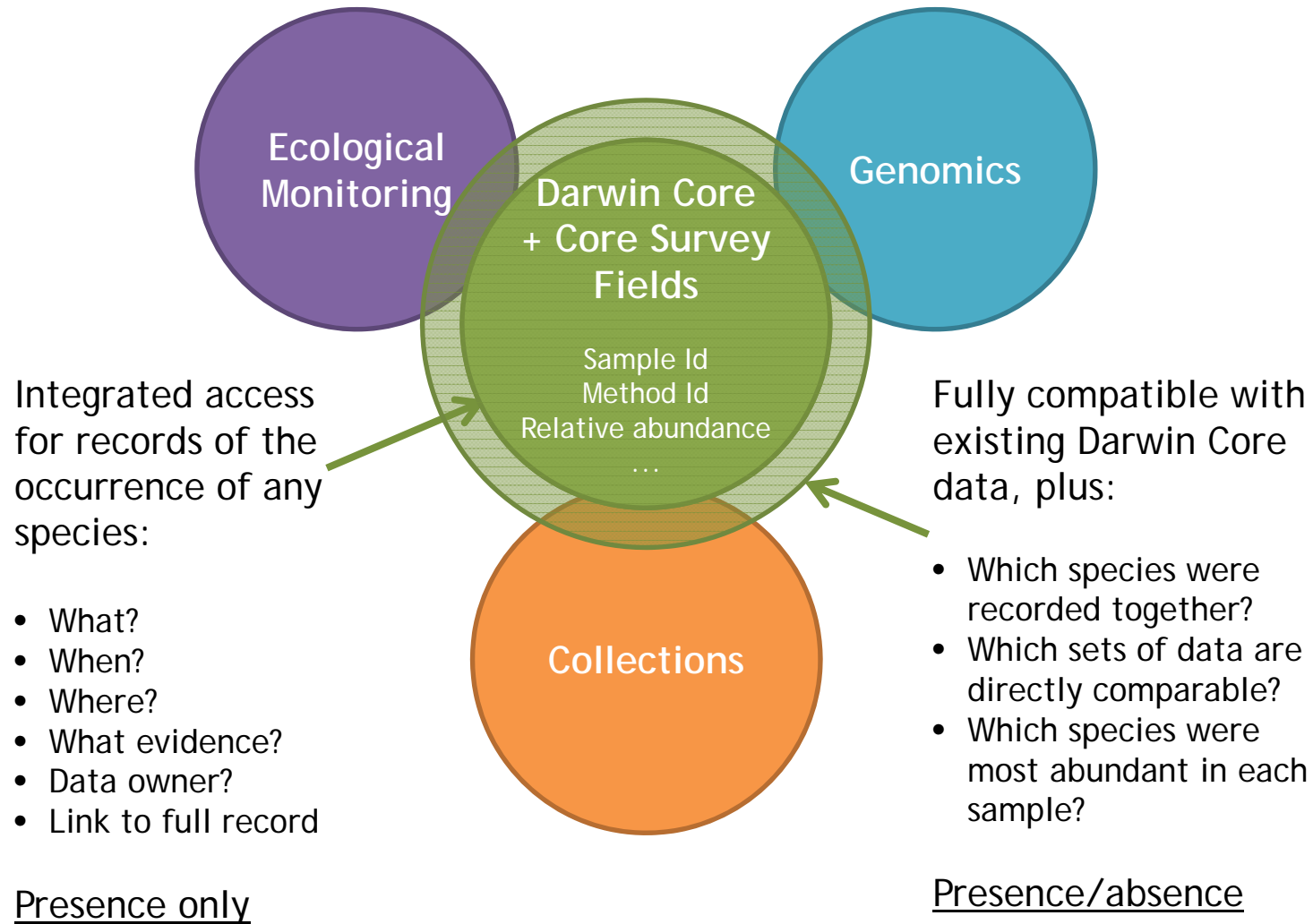
Integrated access
for records of the
occurrence of any
species:

- What?
- When?
- Where?
- What evidence?
- Data owner?
- Link to full record

Presence only



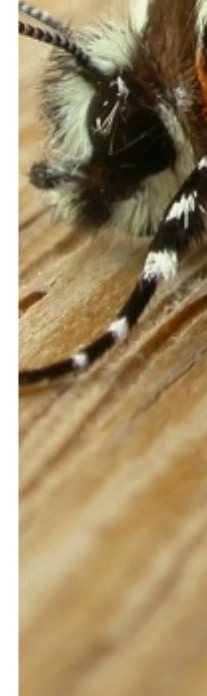
Unifying species data



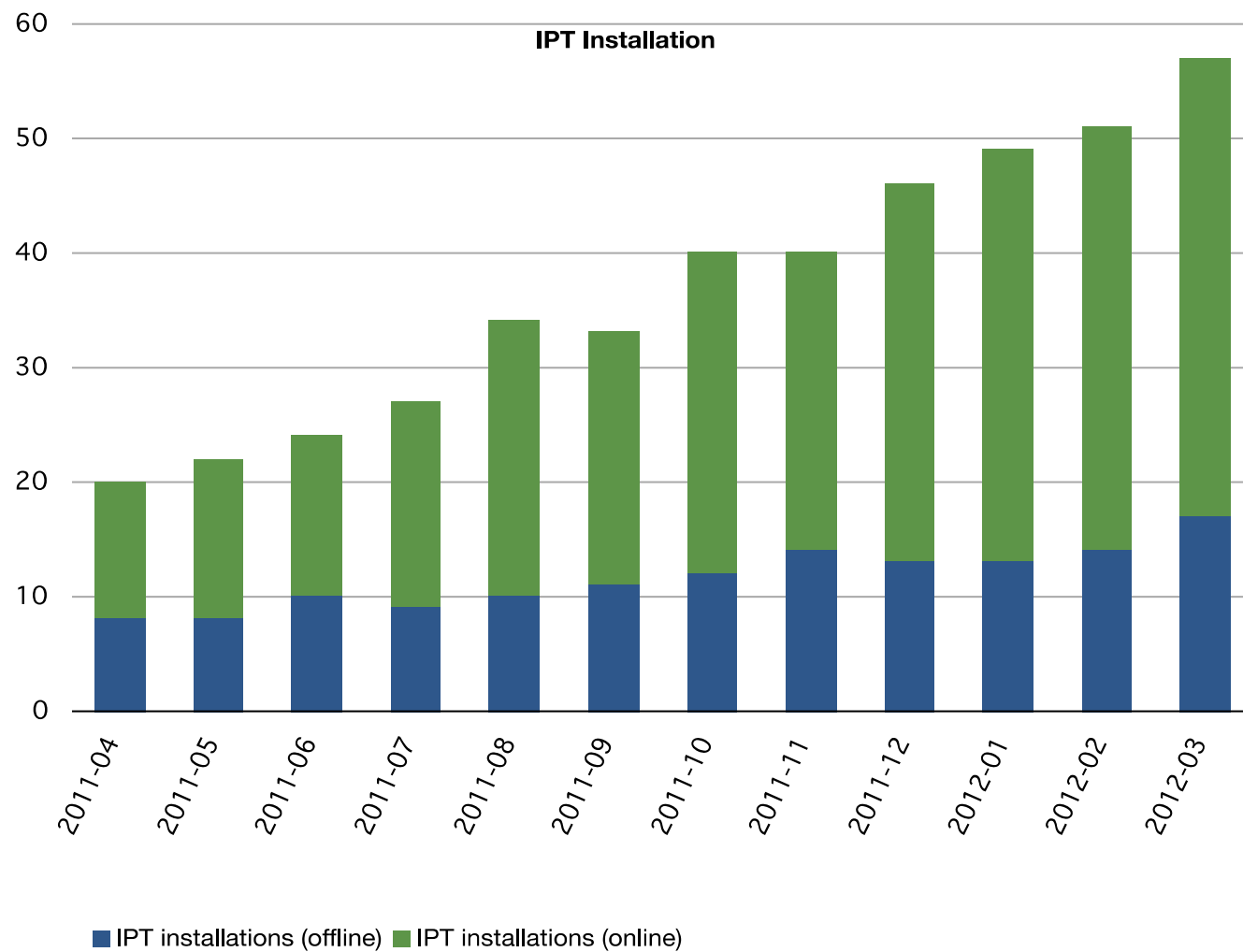
The Integrated Publishing Toolkit (IPT)

A web based tool (Java) enabling:

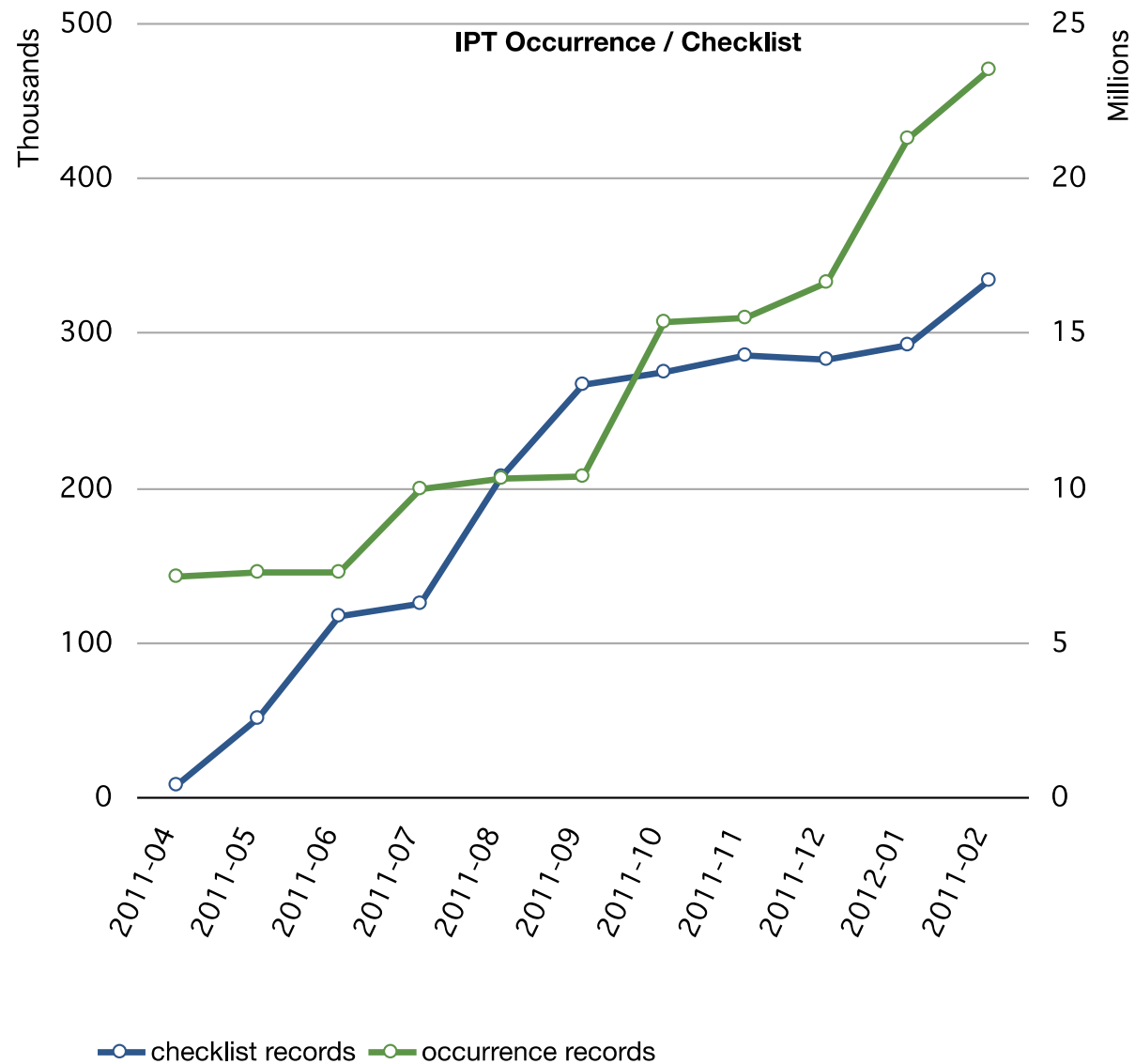
- i. Creation, publication and sharing of DwC-A based data
- i. Single-click registration with GBIF
- ii. Mapping of custom databases to the DwC-A standard
- iii. Uploading of (e.g.) CSV files for mapping to the DwC-A standard
- iv. Authoring of dataset descriptions according to the GBIF metadata profile
- v. Publish data through GBIF, allowing others (e.g. <http://eol.org>) to access same content in standard manner



IPT Uptake



IPT Records Served



IPT related resources

Project website

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

Customizing the style

- <http://www.canadensys.net/2011/customizing-the-ipt>

Demonstration installation

- <http://ipt.gbif.org/>

Software architecture

- http://gbif-providertoolkit.googlecode.com/files/ipt-architecture_1.1.pdf



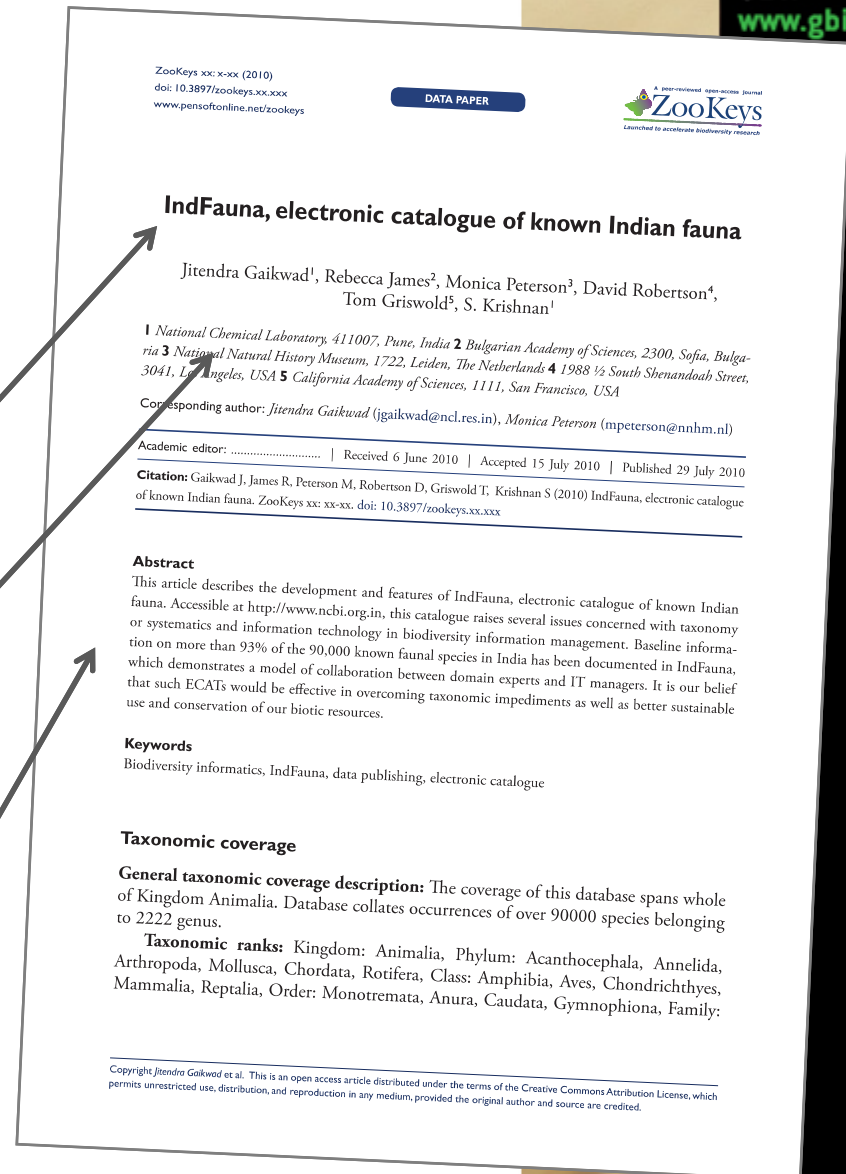
Data Paper

What it is: Scholarly publication of searchable metadata document describing a dataset, or a group of datasets

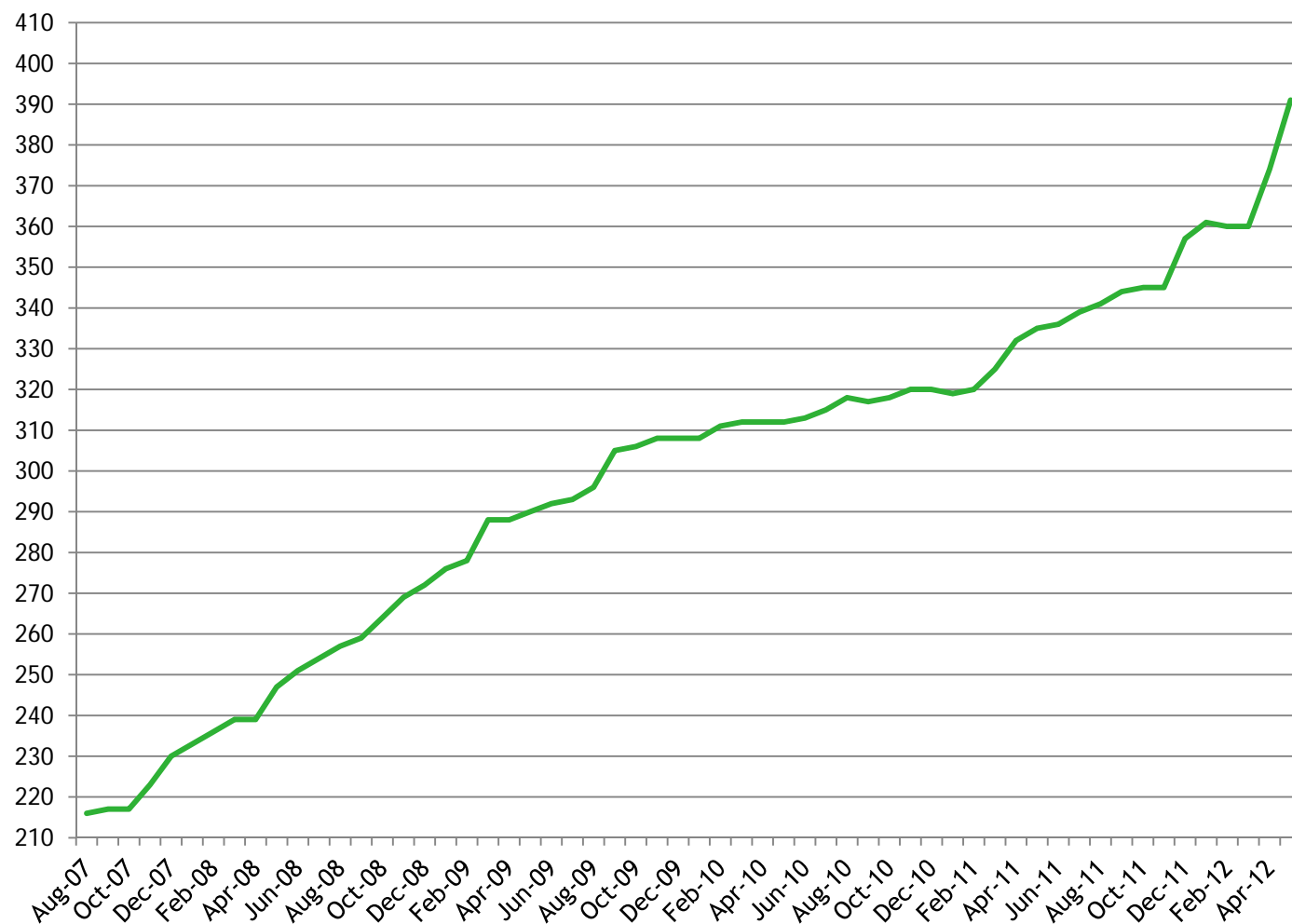
*Promote and publicize
existence of data*

*Provide scholarly credit
to data publishers
through citable journal
publications*

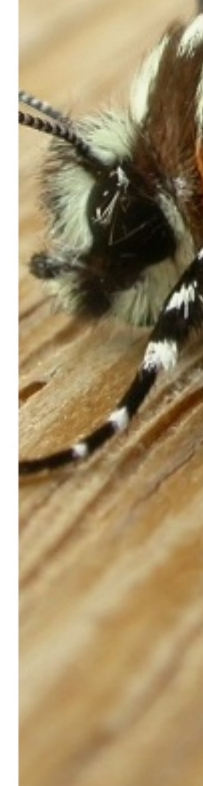
*Describe the data
in a structured
human-readable
form*



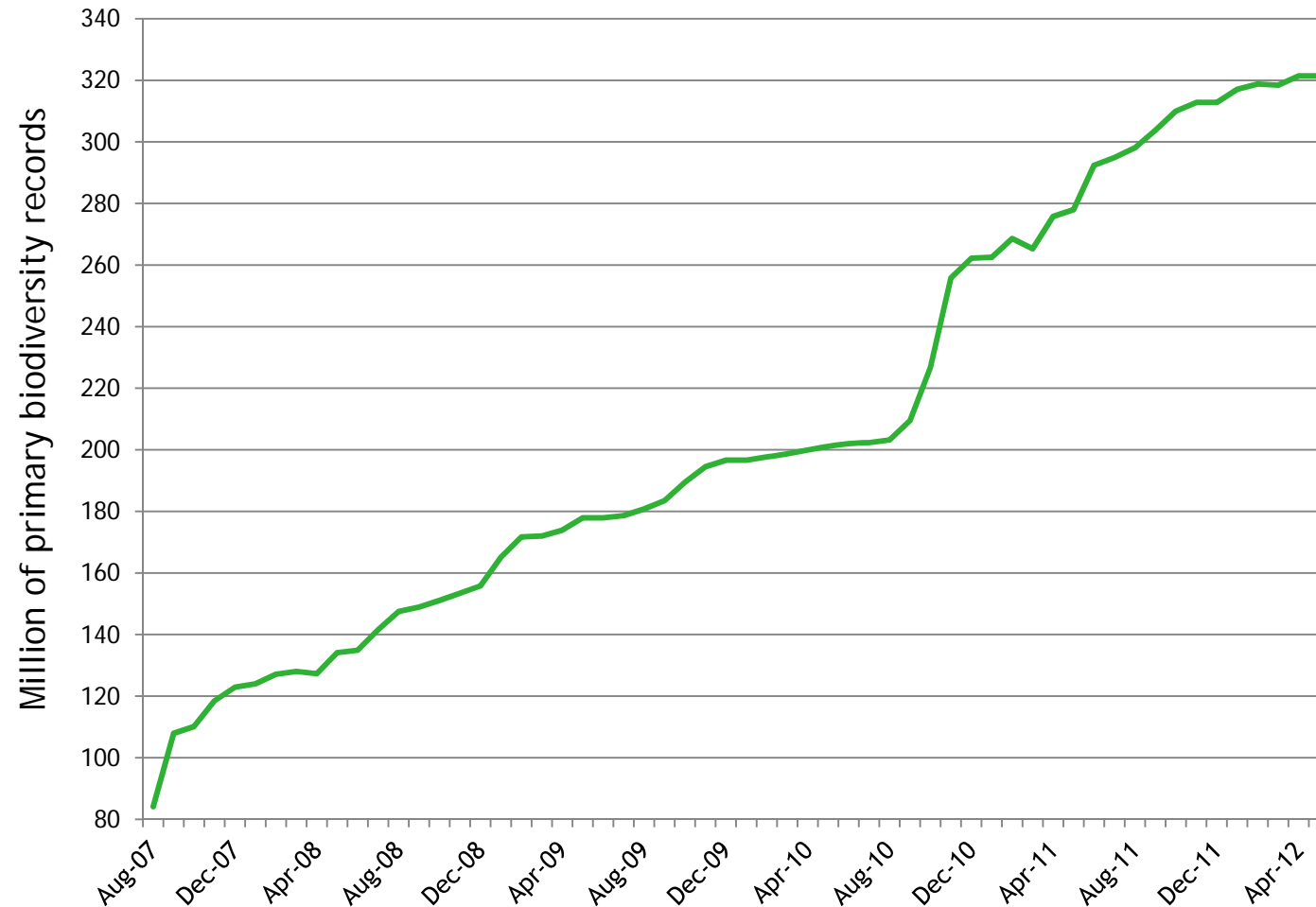
Growth in data publishers



Last updated: 2012-05-02

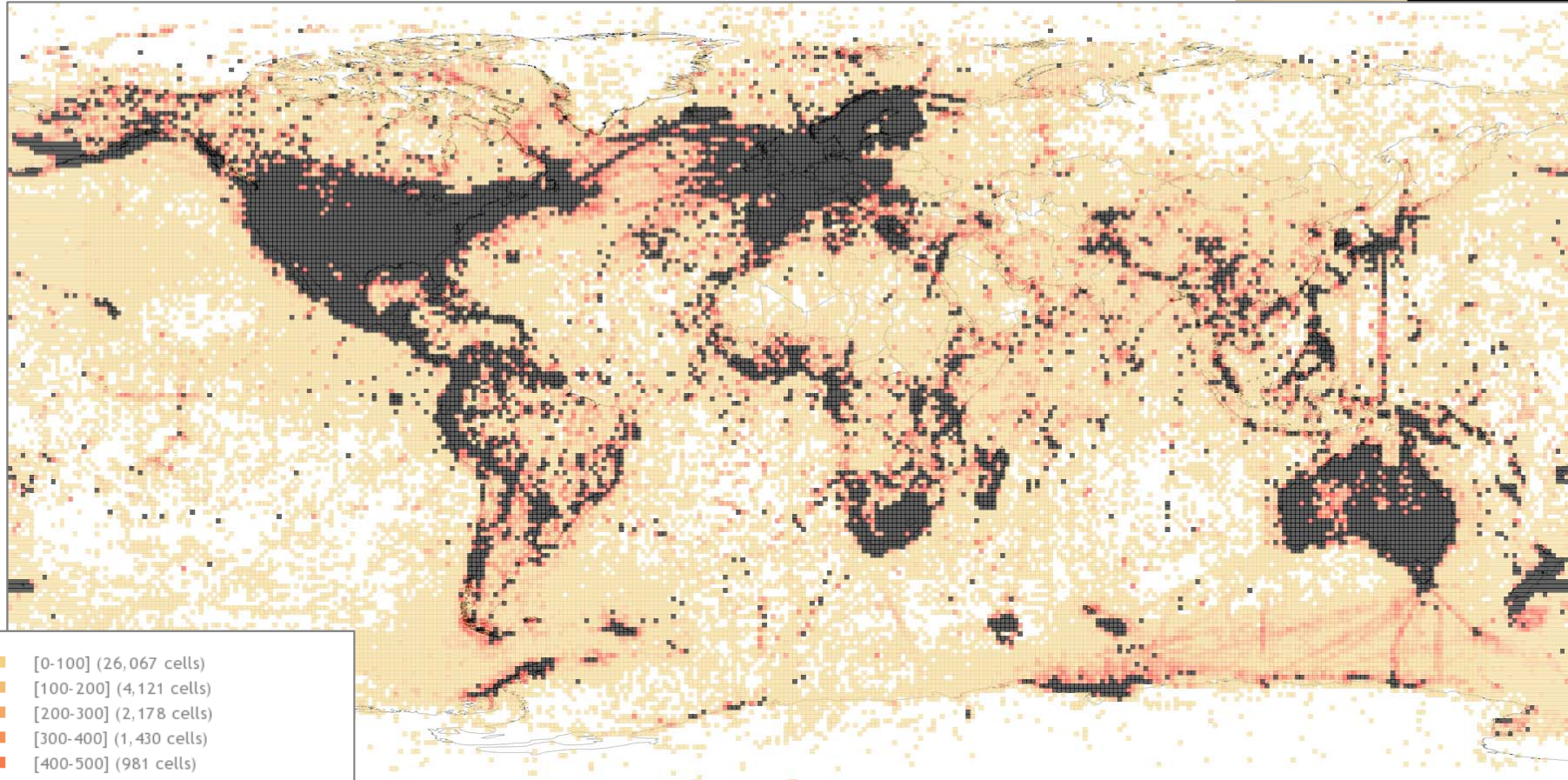


Growth in data records



The coverage in 2012

Occurrences



[0-100]	(26,067 cells)
[100-200]	(4,121 cells)
[200-300]	(2,178 cells)
[300-400]	(1,430 cells)
[400-500]	(981 cells)
[500-600]	(738 cells)
[600-700]	(567 cells)
[700-800]	(437 cells)
[800-900]	(393 cells)
[900-1,000]	(324 cells)
> 1000	(7330 cells)

How much data out there?

The prediction from the network

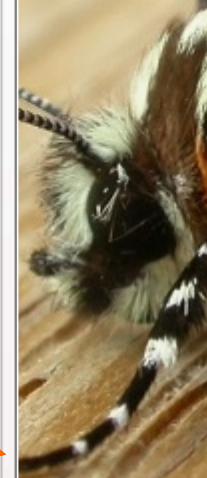
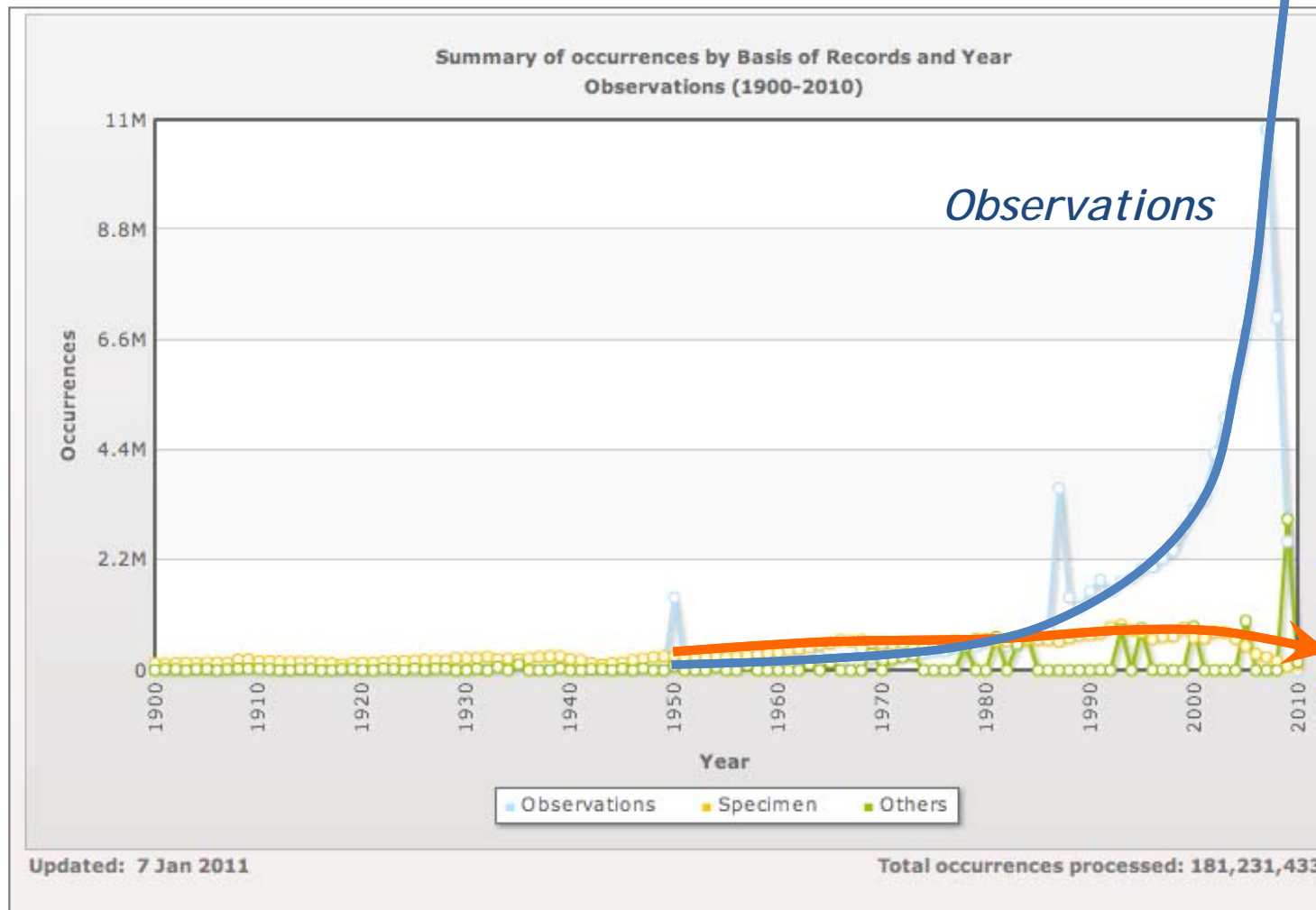
Please provide the best estimate of the TOTAL amount of primary biodiversity data currently (to end 2010) available in your country.			
Record type	Digital	Non-digital	TOTAL
Specimen based occurrence data	135,201,086	1,524,152,550	1,659,353,636
Observation based occurrence records	610,227,574	441,170,000	1,051,397,574
Multimedia data linked to primary biodiversity data	5,120,671	2,000,000	7,120,671
Population / ecological monitoring records	88,012,225	0	88,012,225
Impact Assessment associated data records	1,505,000	15,000	1,520,000
Other types of primary biodiversity data	100,095,000	0	100,095,000
TOTAL	940,161,556	1,967,337,560	2,907,499,106
<small>N=31 countries Source: GBIF Participant Report 2010</small>			

- 19 Country Participants estimated being able to publish a total of 471.67 million records by end 2016
- Five Organisation Participants estimated being able to publish a total of 2.7 million records by the end of 2016



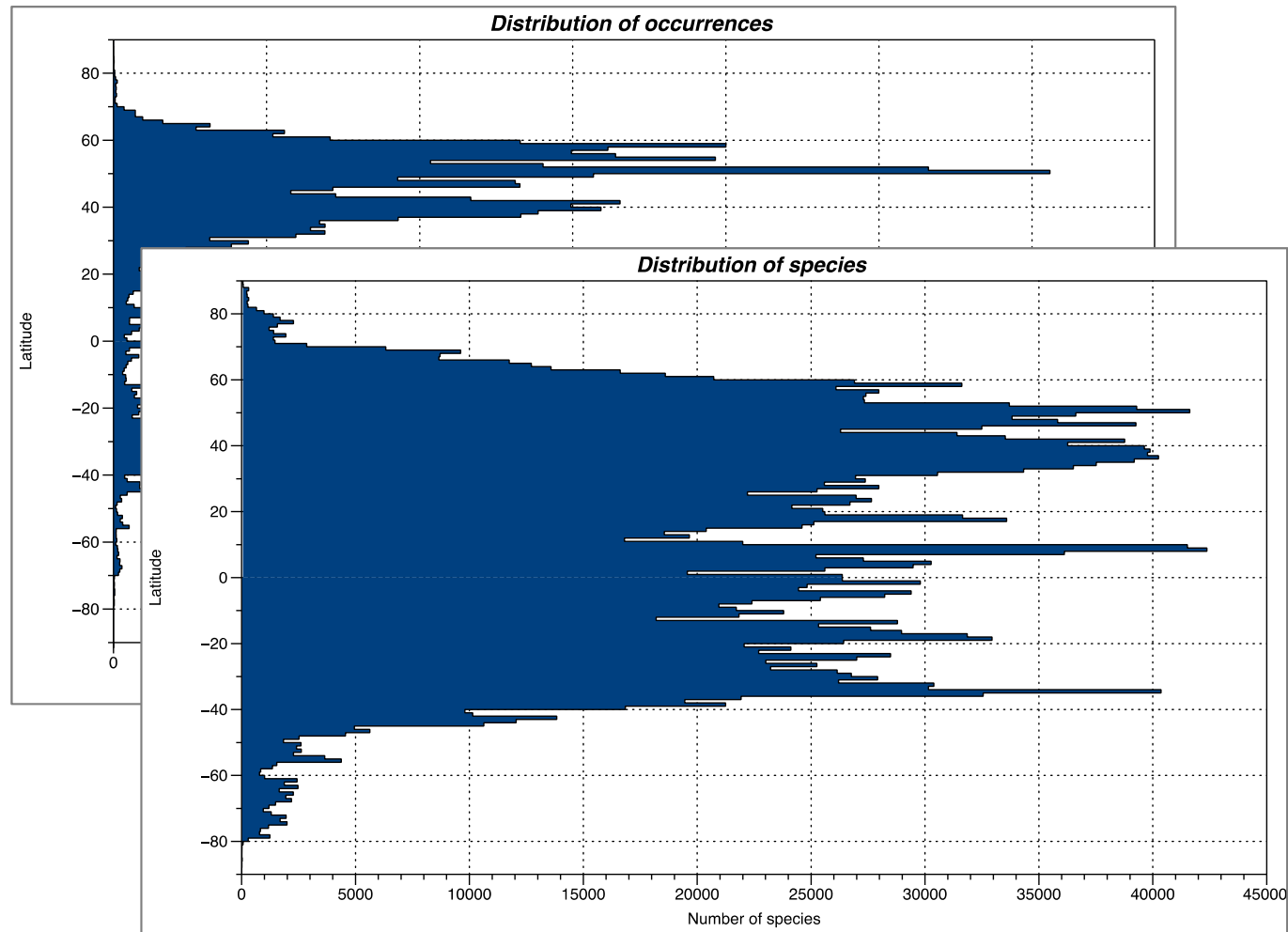
Trends

The growth of observation data

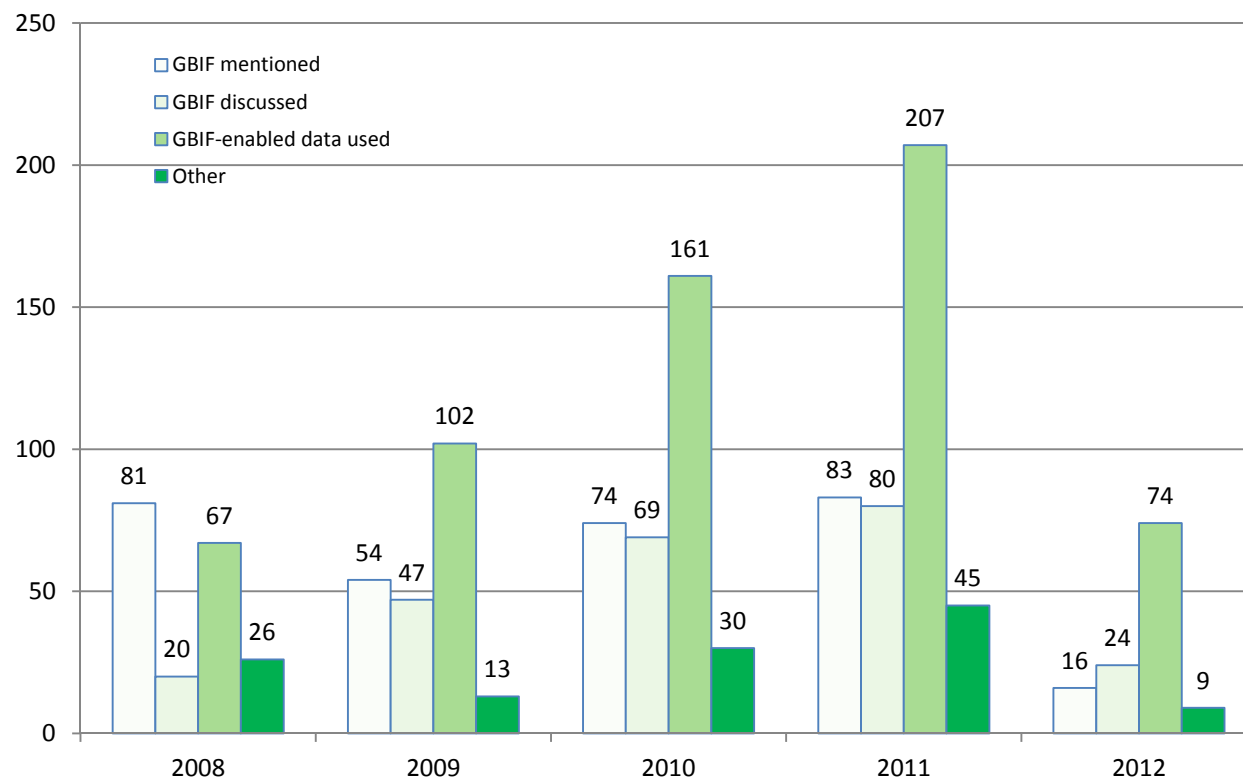


The coverage in 2012

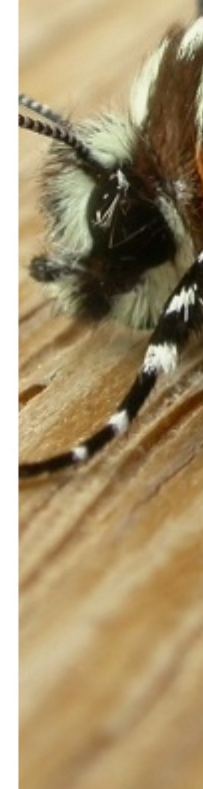
Occurrences/Species



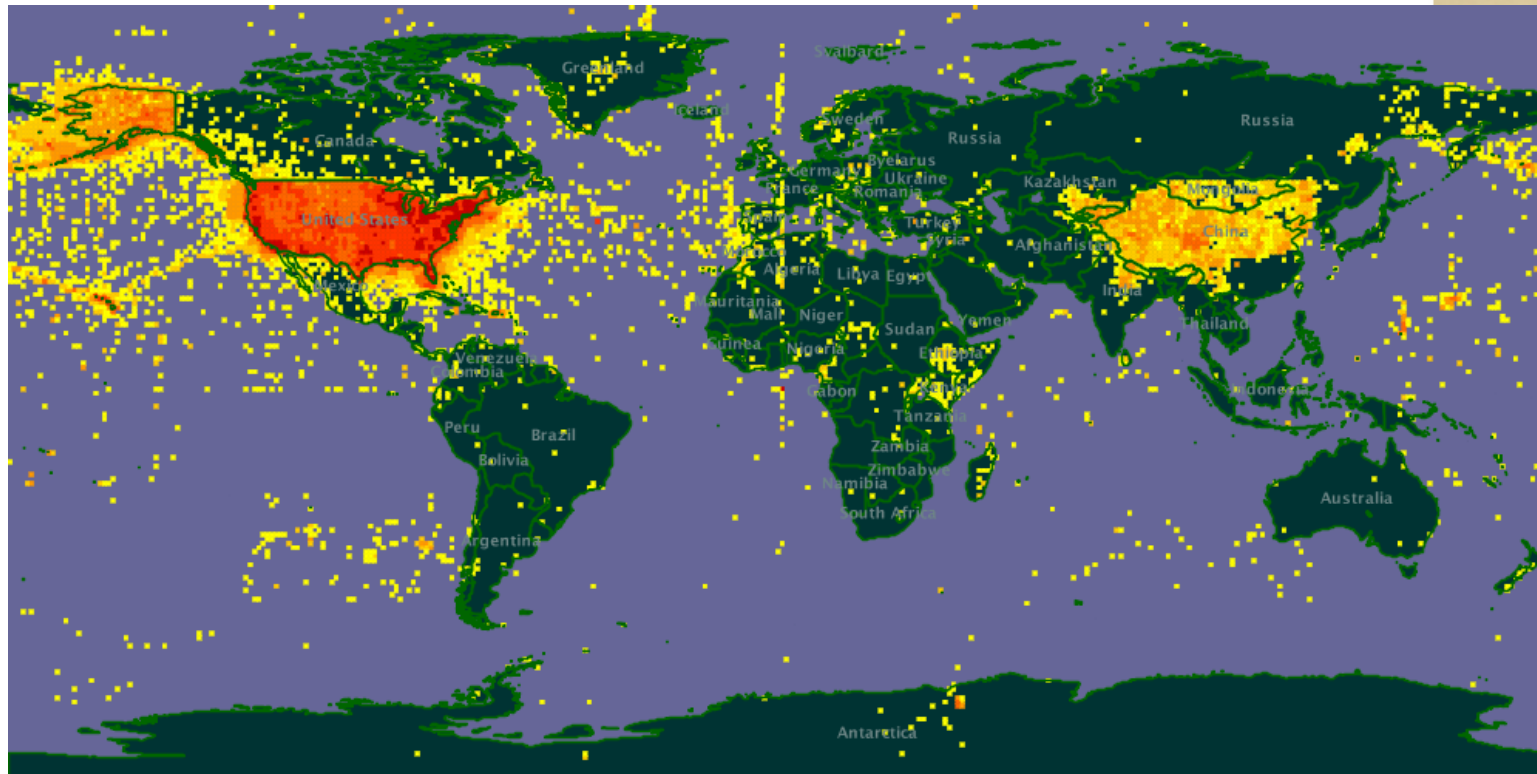
GBIF: Professional publications



	GBIF mentioned	GBIF discussed	GBIF-enabled data used	Other	Total
2008	81	20	67	26	194
2009	54	47	102	13	216
2010	74	69	161	30	334
2011	83	80	207	45	415
2012	16	24	74	9	123
Total	308	240	611	123	1282

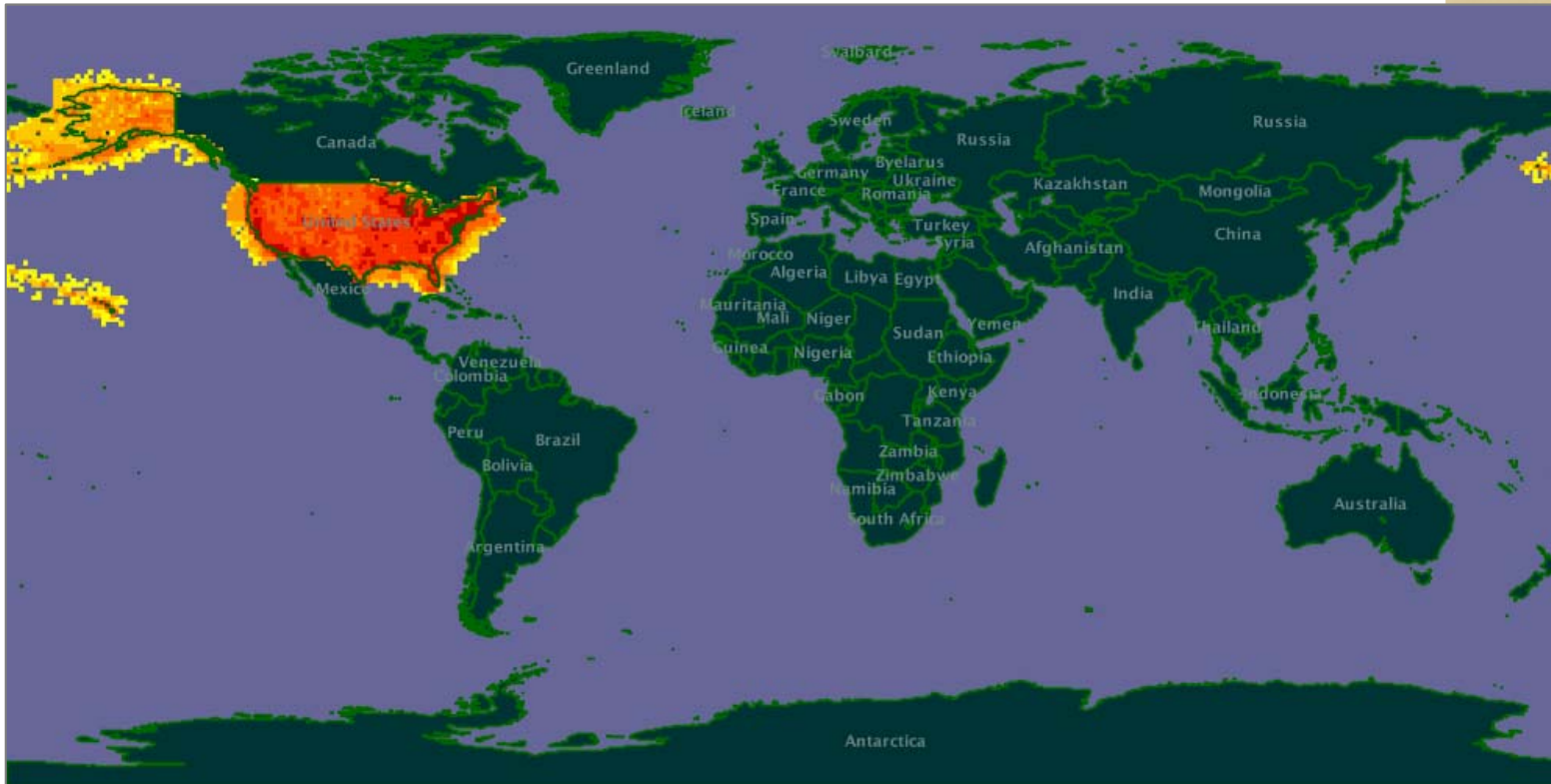


Challenges in data quality



Records with coordinates, claiming to arise from US (or with a country string we are able to interpret as US)

Challenges in data quality



Following the new Portal processing

ChecklistBank

The component of the GBIF architecture related to taxonomic organization

- i. By communicating with the GBIF registry, discovers and indexes checklists published through GBIF
- i. “Harmonizes” data into a consistent format
 - Stored in a PostGRES database
- ii. Supports the DwC-A input format
- iii. Integrates authoritative checklists into a pseudo-taxonomy used for the “backbone” of the GBIF data portal (<http://data.gbif.org>)
- iv. Offers a set of web services to assist in (e.g.) synonym resolution and name linkages
- v. Progressively being incorporated into the GBIF Data Portal to enrich taxonomic content within the portal



ChecklistBank Prototype web app.

<http://ecat-dev.gbif.org>

The screenshot displays the GBIF Checklist Bank web application interface. The browser window title is "GBIF Backbone Taxonomy - Lycopodiaceae Palisot de Beauvois ex de Mirbel, 1802". The address bar shows the URL "http://ecat-dev.gbif.org/usage/6725". The page features a navigation bar with "Checklists", "API", "Tools", and "About" links. A "BETA" stamp is visible. A search bar is present with the text "search GBIF Backbone Taxonomy". The main content area is titled "Lycopodiaceae Palisot de Beauvois ex de Mirbel, 1802". It includes a "GBIF Backbone Taxonomy" section with details on the source (The Catalogue of Life, 3rd January 2011) and publication (J. B. A. P. M. de Lamarck & C. F. B. de Mirbel, Hist. Nat. Vég. 4: 293. 21 Nov 1802). A "Classification" sidebar lists the taxonomic hierarchy: Kingdom Plantae, Phylum Lycopodiophyta, Class Lycopodiopsida, Order Lycopodiales, Family Lycopodiaceae, Subfamily ---, Genus 26 usages, Subgenus ---, and Species 1,379 usages. A "Synonyms" section lists "Selago J. Hill, 1757" as a genus synonym. A "Children" section lists various genera including Austrolycopodium, Copodium, Dendrolycopodium, Diphasiastrum, Diphasium, Huperzia, Lateristachys, Lepidotis, Lycopodiastrium, Lycopodiella, Lycopodiopsis, Lycopodites, Lycopodium, Lycoxylon, and Mesostrobilus. A "Nub Disclaimer" section explains that the information is automatically synthesized and may conflict with other data. An "Identifier" section is also present.

GBIF Checklist Bank **BETA**

Checklists API Tools About

search GBIF Backbone Taxonomy Search

Indexed Registered ToC

Lycopodiaceae Palisot de Beauvois ex de Mirbel, 1802

GBIF Backbone Taxonomy taxonomic

AccordingTo The Catalogue of Life, 3rd January 2011

PublishedIn in J. B. A. P. M. de Lamarck & C. F. B. de Mirbel, Hist. Nat. Vég. 4: 293. 21 Nov 1802.

TaxonStatus accepted NomenStatus --- Rank family

Synonyms hide

Selago J. Hill, 1757 genus synonym

Children hide

Austrolycopodium Holub genus [8]
Copodium Rafinesque, 1817 genus
Dendrolycopodium A. Haines, 1993 genus [4]
Diphasiastrum Holub genus [22]
Diphasium Rothm. genus [18]
Huperzia Bernh. genus [537]
Lateristachys Holub genus [4]
Lepidotis Mirb. genus [6]
Lycopodiastrium J. Holub genus [1]
Lycopodiella Holub genus [90]
Lycopodiopsis Renault, 1890 genus
Lycopodites Lindley & W. Hutton, 1833 genus [6]
Lycopodium L. genus [710]
Lycoxylon B.P. Srivastava, 1944 genus
Mesostrobilus J. Watson, 1909 genus

Classification

Direct Lycopodiales de Candolle ex
Parent Berchtold & J. Presl

Kingdom Plantae
Phylum Lycopodiophyta
Class Lycopodiopsida
Order Lycopodiales
Family Lycopodiaceae
Subfamily ---
Genus 26 usages show
Subgenus ---
Species 1,379 usages show

Nub Disclaimer

The GBIF Backbone Taxonomy (Nub) is an automatically synthesised management classification with limited manual curating. Information presented here does not represent a consistent taxon but may conflict with other nub "usages" in many cases to a trained taxonomists eye. The information presented on this page was aggregated from the data found in the sources below.

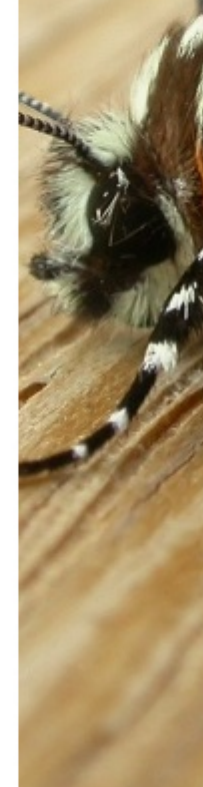
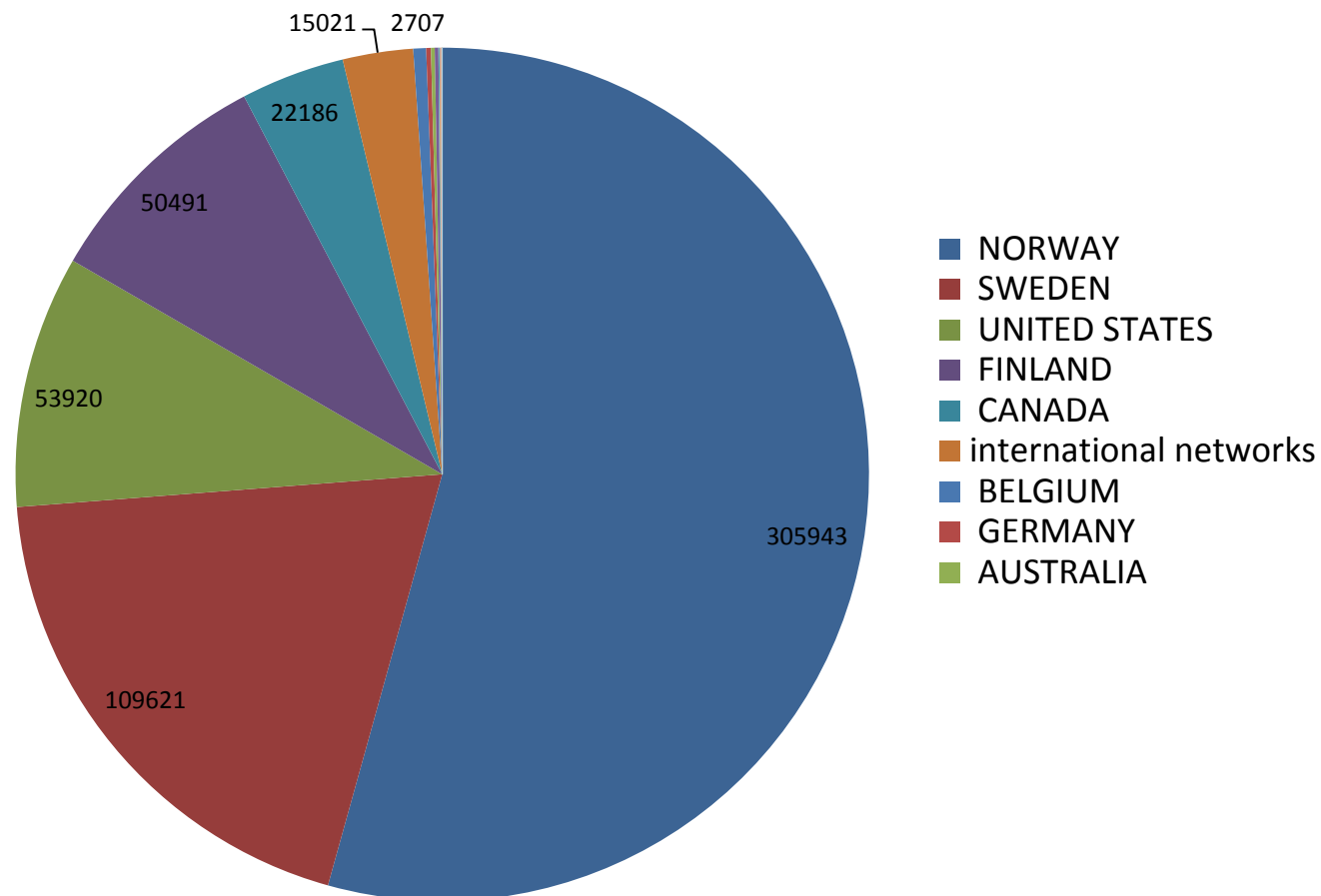
Identifier

Arctic Plant Occurrence Data

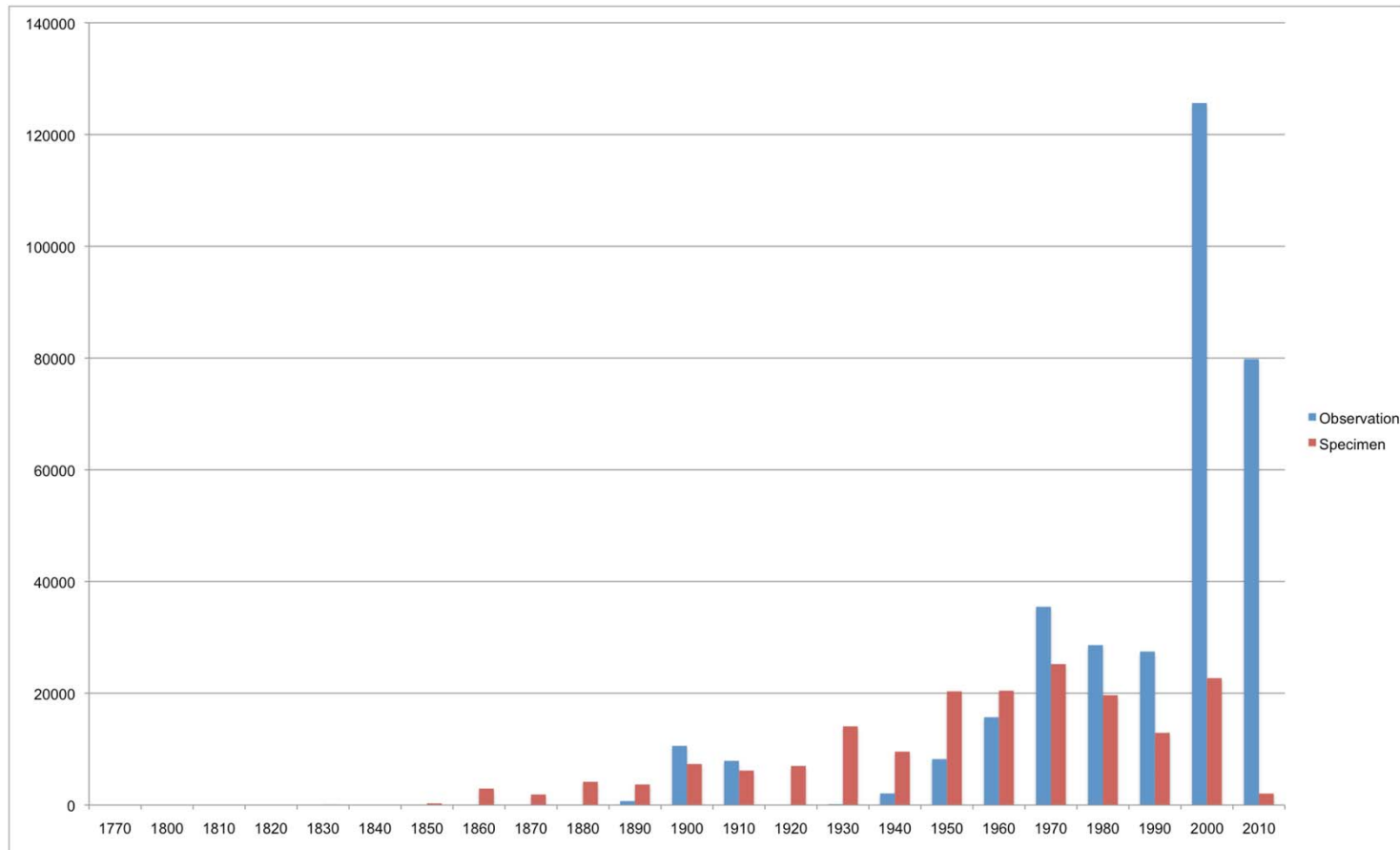
- 1,899,339 georeferenced occurrence records from above the Arctic circle are available through GBIF, 563,274 of them for plants
- The plant records are published by 22 countries, islands and territories and by international organisations through 86 individual institutions



Countries Publishing Arctic Plant Occurrence Data



Record basis by decade (Arctic plant occurrences)



Challenges and Options

- Name resolution PAF/ GBIF global names architecture
 - ([The Species 2000 and ITIS Catalogue of Life: The Catalogue of Life](#))
- Publish PAF as checklist in GBIF
 - Enable mapping of data between IAVD to GBIF
 - Enable publication of data papers
 - Enriched checklist data feed into e.g. EOL
- Integrate arctic data into global data pool
 - Development of quantitative data in GBIF
 - Historic data
 - Influence GBIF development

