



## Sharing and using biodiversity data for e-Science

Cees H.J. Hof

Netherlands Biodiversity Information Facility (NLBIF)

The Dutch branch of the Global Biodiversity Information Facility (GBIF)

## What is GBIF?

The Global Biodiversity Information Facility (GBIF) is an **international open data infrastructure**, funded by governments.

It allows **anyone, anywhere to access data** about all types of life on Earth, shared across national boundaries via the Internet.

By encouraging and helping institutions to publish data according to **common standards**, GBIF enables **research** not possible before, and informs better decisions to conserve and sustainably use the biological resources of the planet.

GBIF operates through **a network of nodes**, coordinating the biodiversity information facilities of Participant countries and organizations, collaborating with each other and the Secretariat to share skills, experiences and technical capacity.

**GBIF's vision:** "A world in which biodiversity information is freely and universally available for science, society and a sustainable future."



## History



1999

The purpose of GBIF is to co-ordinate the **standardisation**, **digitisation** and global **dissemination** (within an appropriate property rights framework) of the world's biodiversity data.

**Megascience Forum  
Working Group on  
Biological Informatics**

Ministers acknowledged the importance of such a facility in the areas of **health, resource management, environmental protection, agriculture and education**



2001

GBIF established and secretariat settled in Copenhagen



**NLBIF**

2002

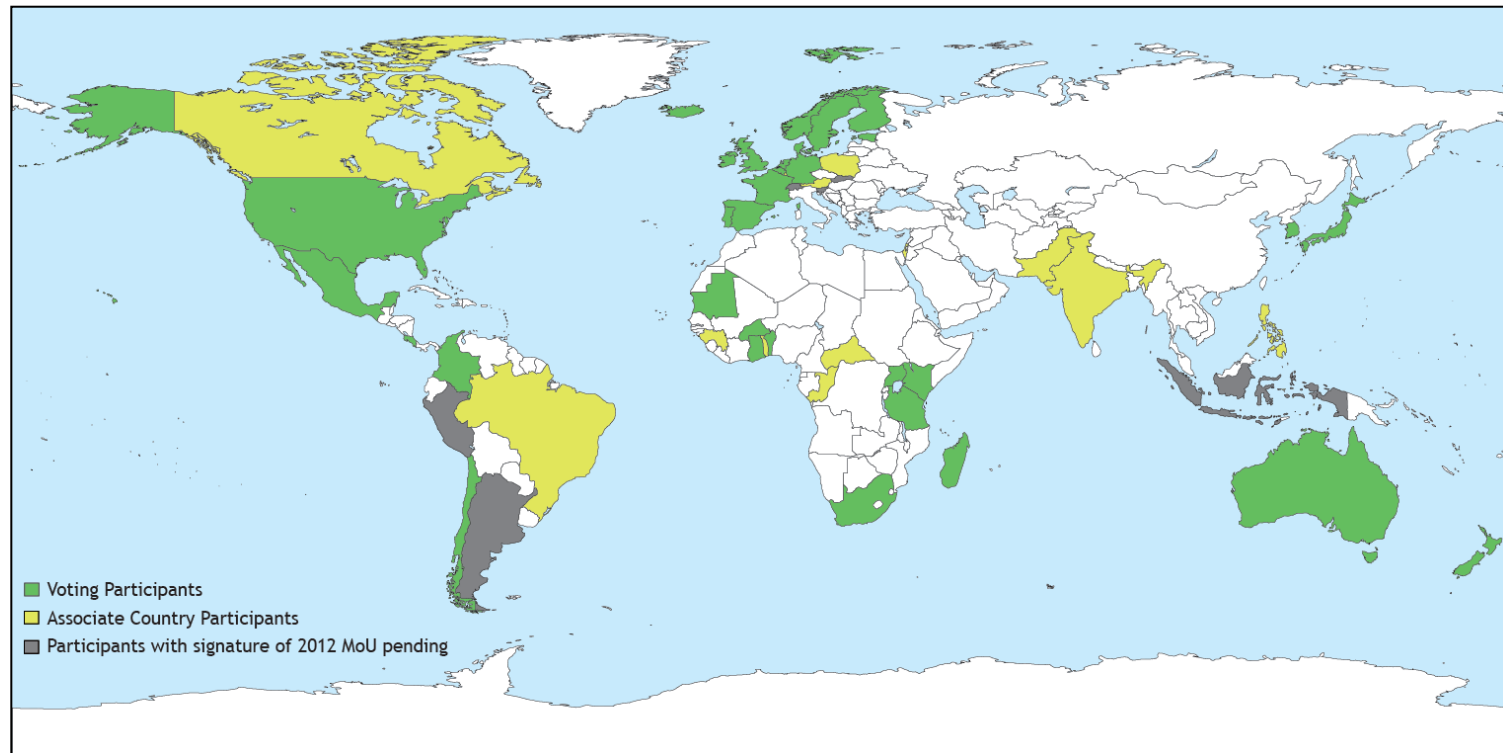
Founded by OCW, NWO and UvA

2012

Currently **structurally** financed by OCW, hosted by UvA

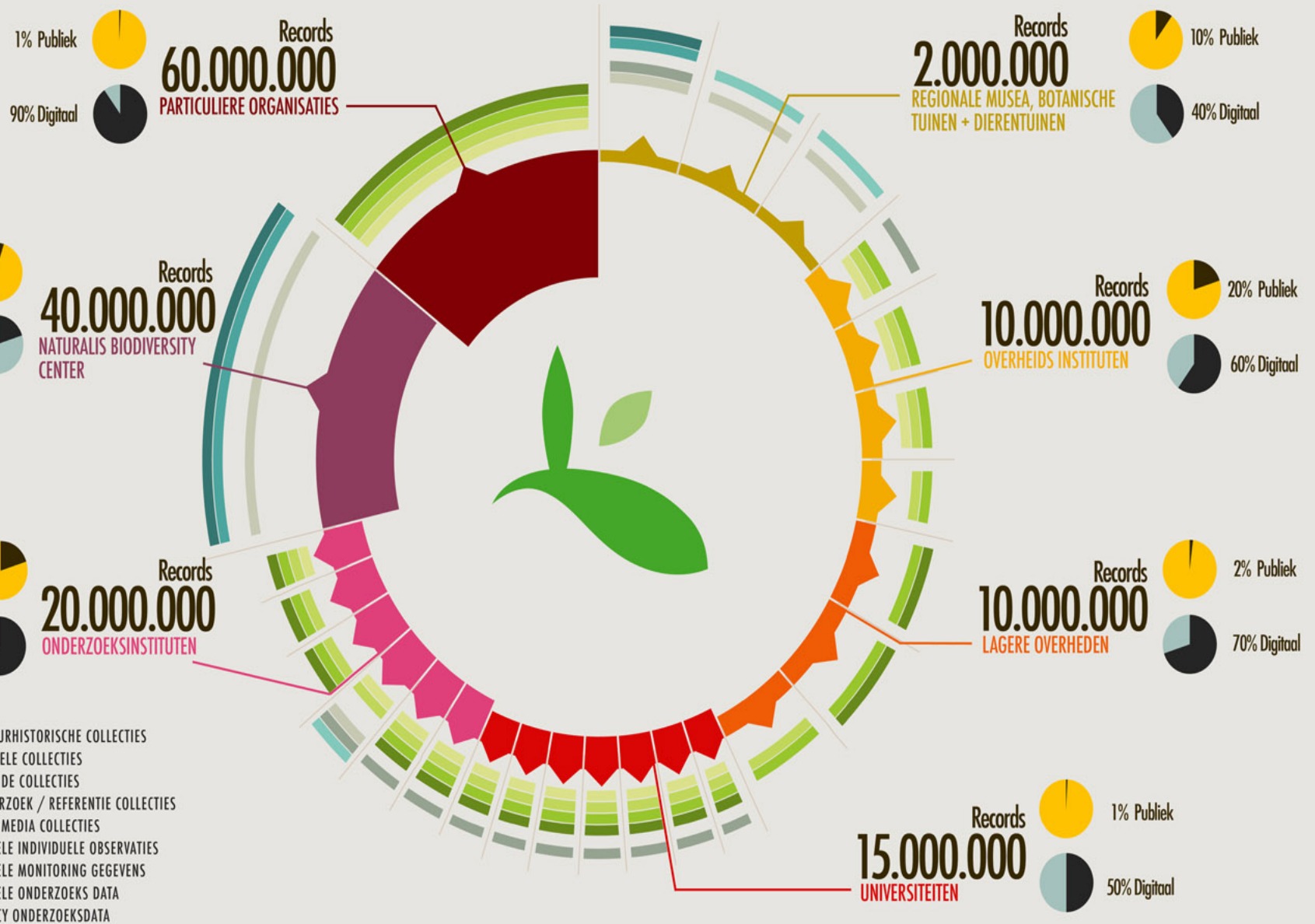


# GBIF country participants

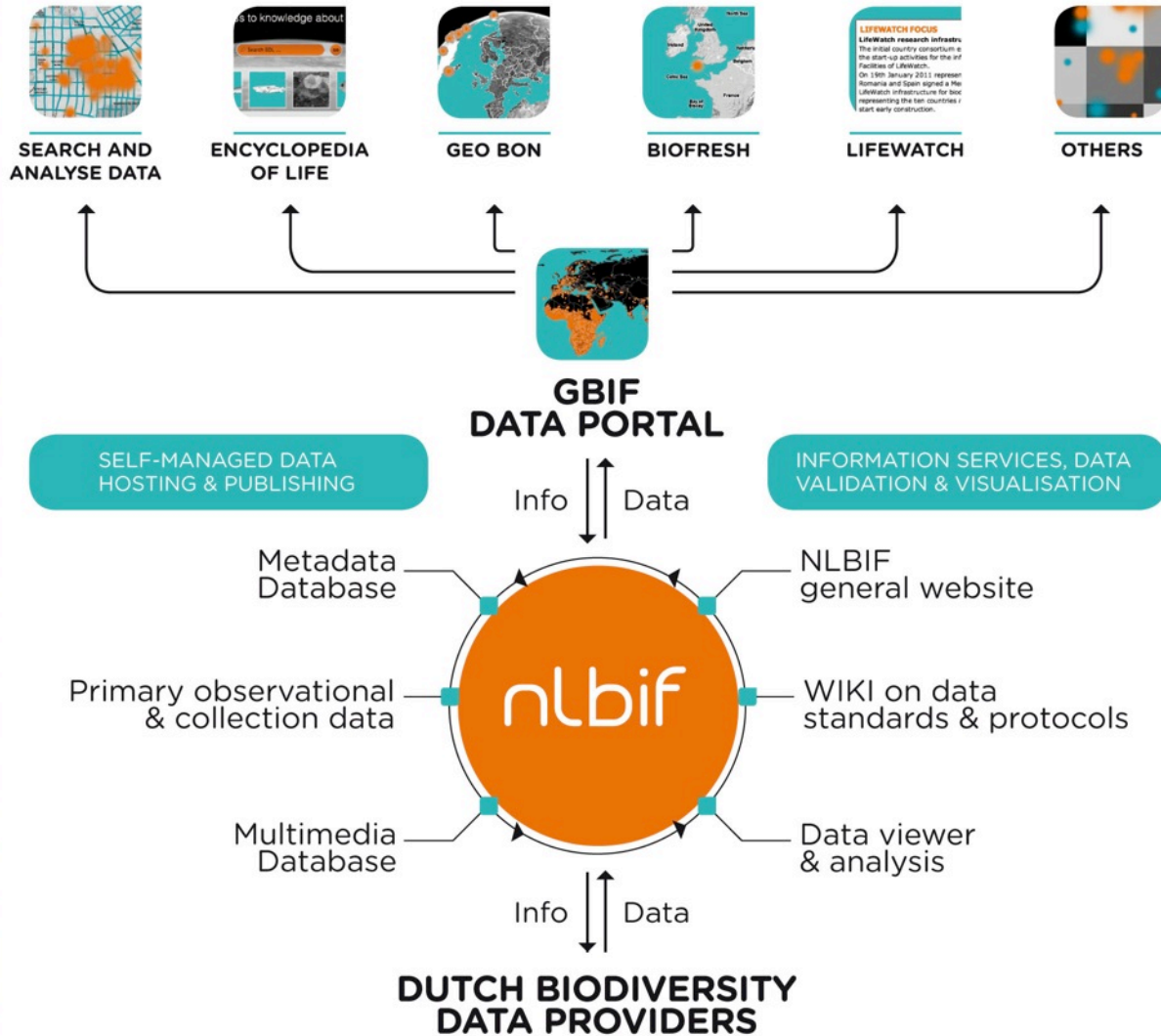


# DATA TYPE

- Natuurhistorische collecties
- Fossiele collecties
- Levende collecties
- Onderzoek / referentie collecties
- Multimedia collecties
- Actuele individuele observaties
- Actuele monitoring gegevens
- Actuele onderzoeks data
- Legacy onderzoeksdata



# Netherlands Biodiversity Information Facility



GBIF data flow:

presentation

indexing

publishing

registration  
(+ archiving)

standardisation

data generation

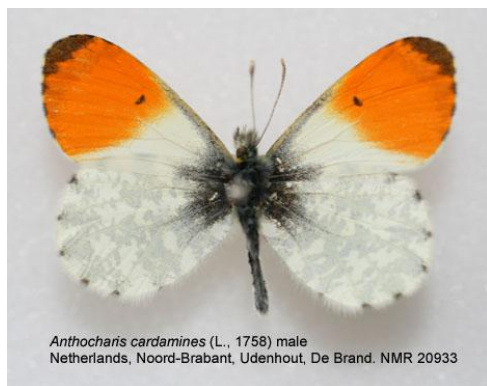
## The data:

What is a record in GBIF?

GBIF hosts so called “occurrence data”

→ Which species of plant / animal / micro-organism was observed / collected from which locality and when?

Taxonomic information



Geographic information

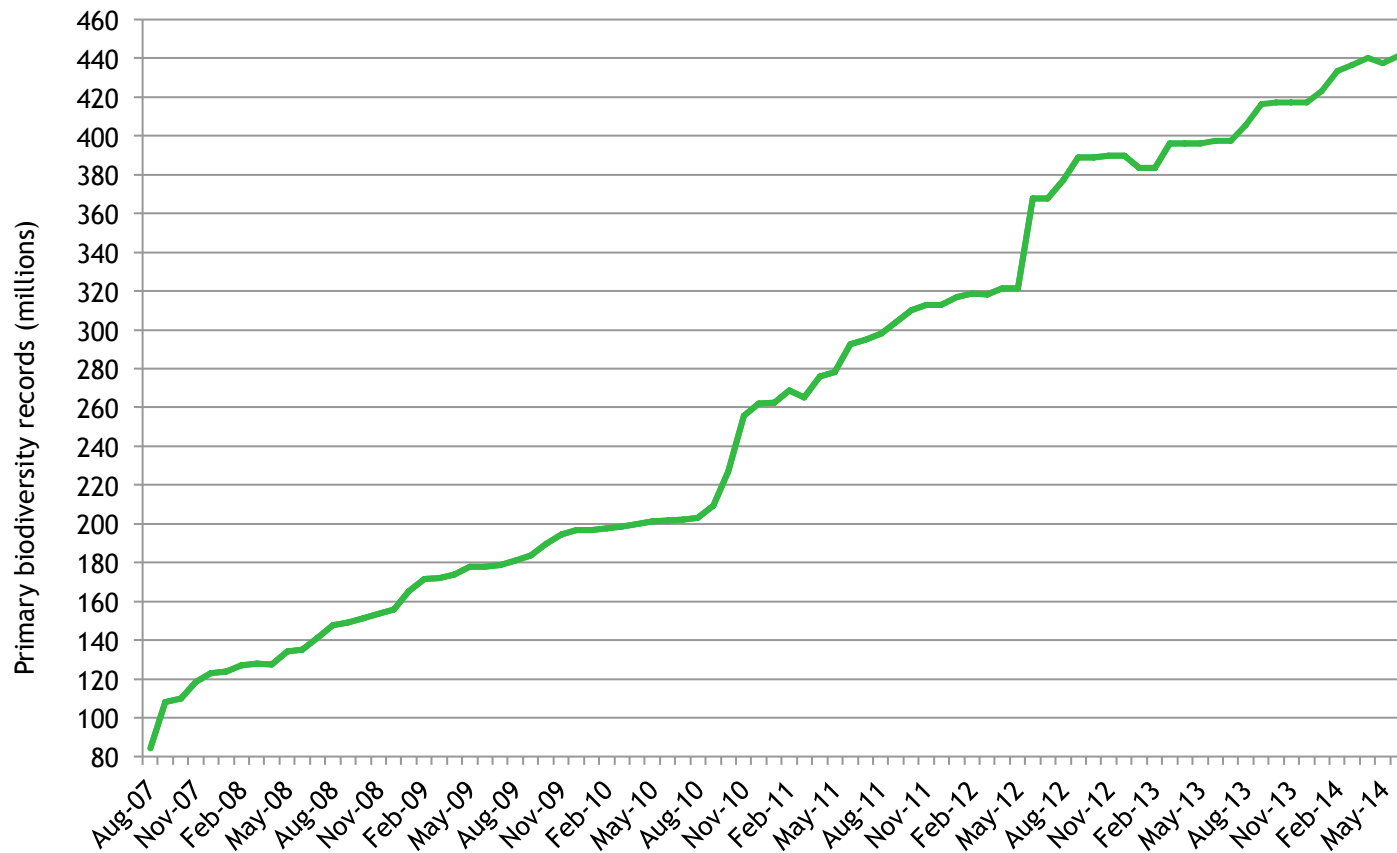


Temporal information

01 May 1971  
00:00:00

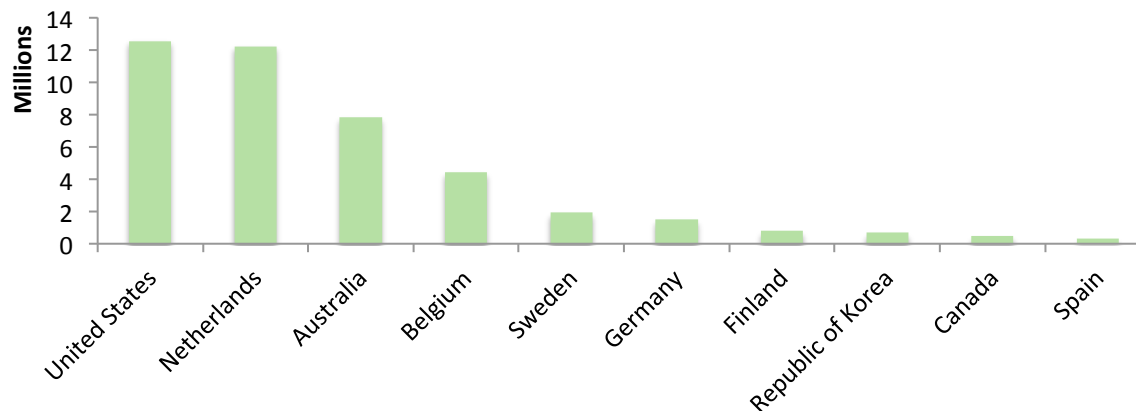


# Data published through GBIF






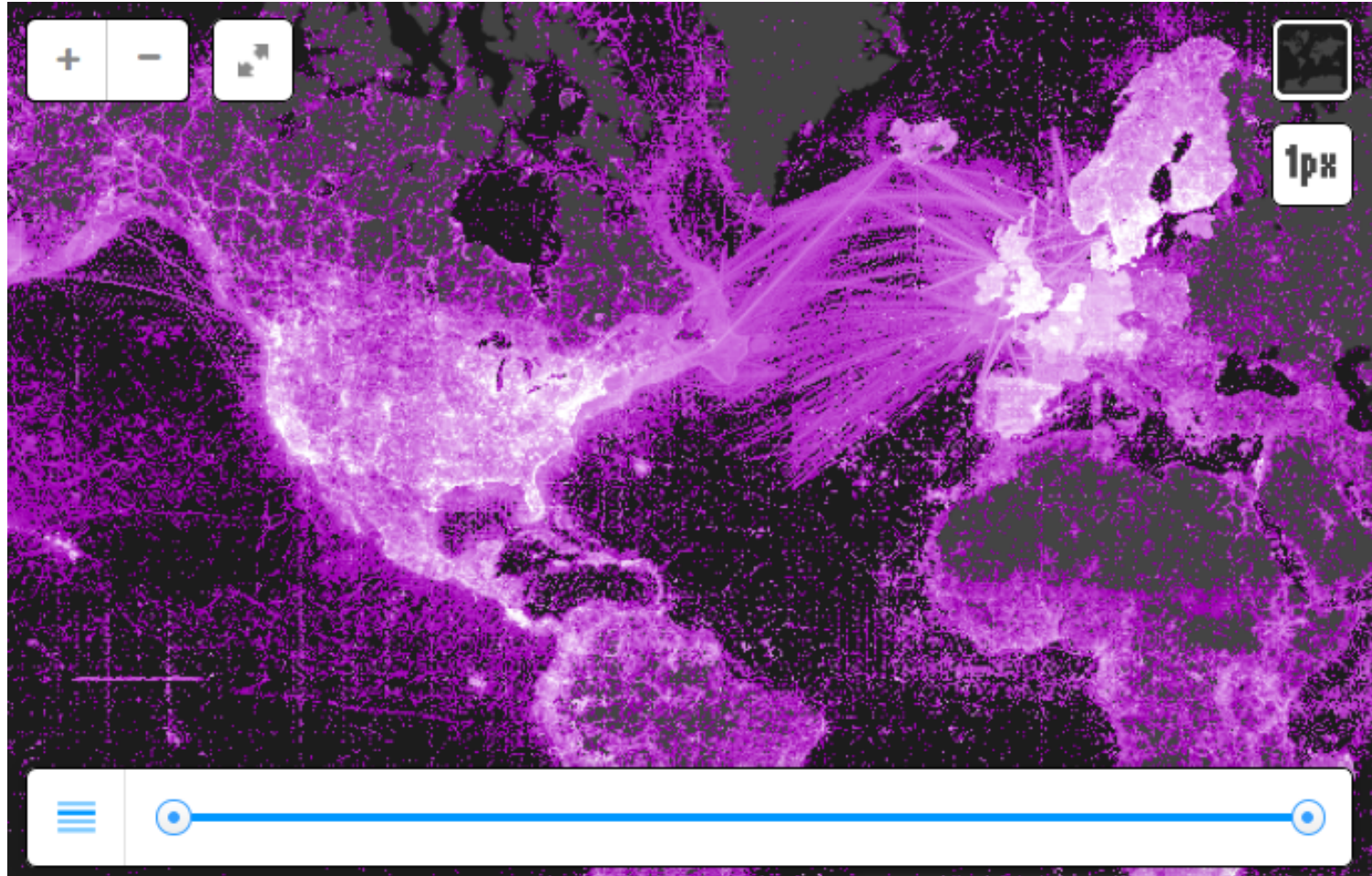
### Top 10 Participant Countries—Number of new records published (1 January - 31 May 2014)



### Top 10 Participant Countries—Total number of records published (as of 31 May 2014)

Country	Records published (in millions)
 United States	159.6
 Sweden	46.1
 United Kingdom	42.4
 Australia	32.3
 Netherlands	18.0
 Finland	16.8
 France	16.4
 Germany	13.5
 Norway	13.0
 Denmark	9.3



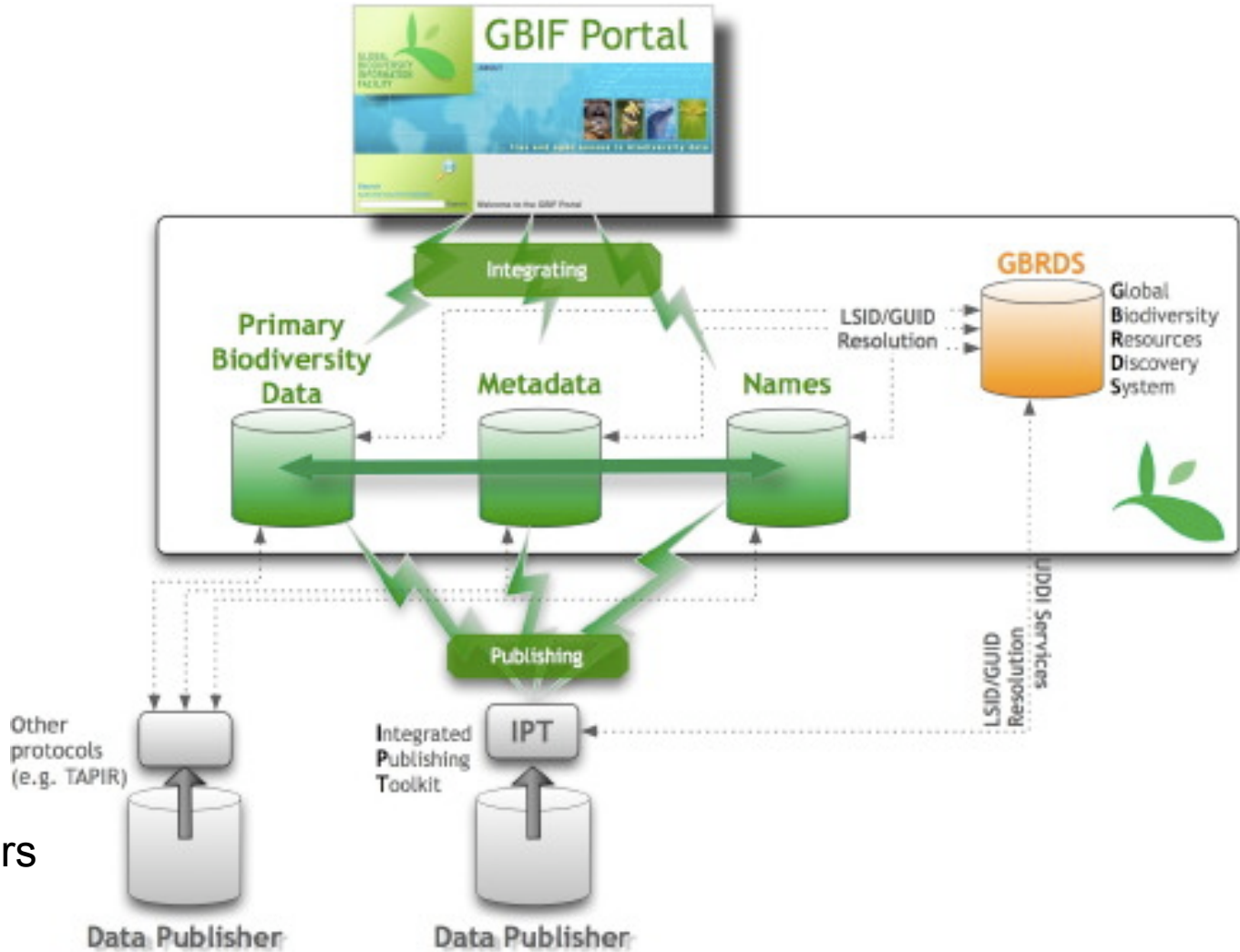


# The GBIF infrastructure

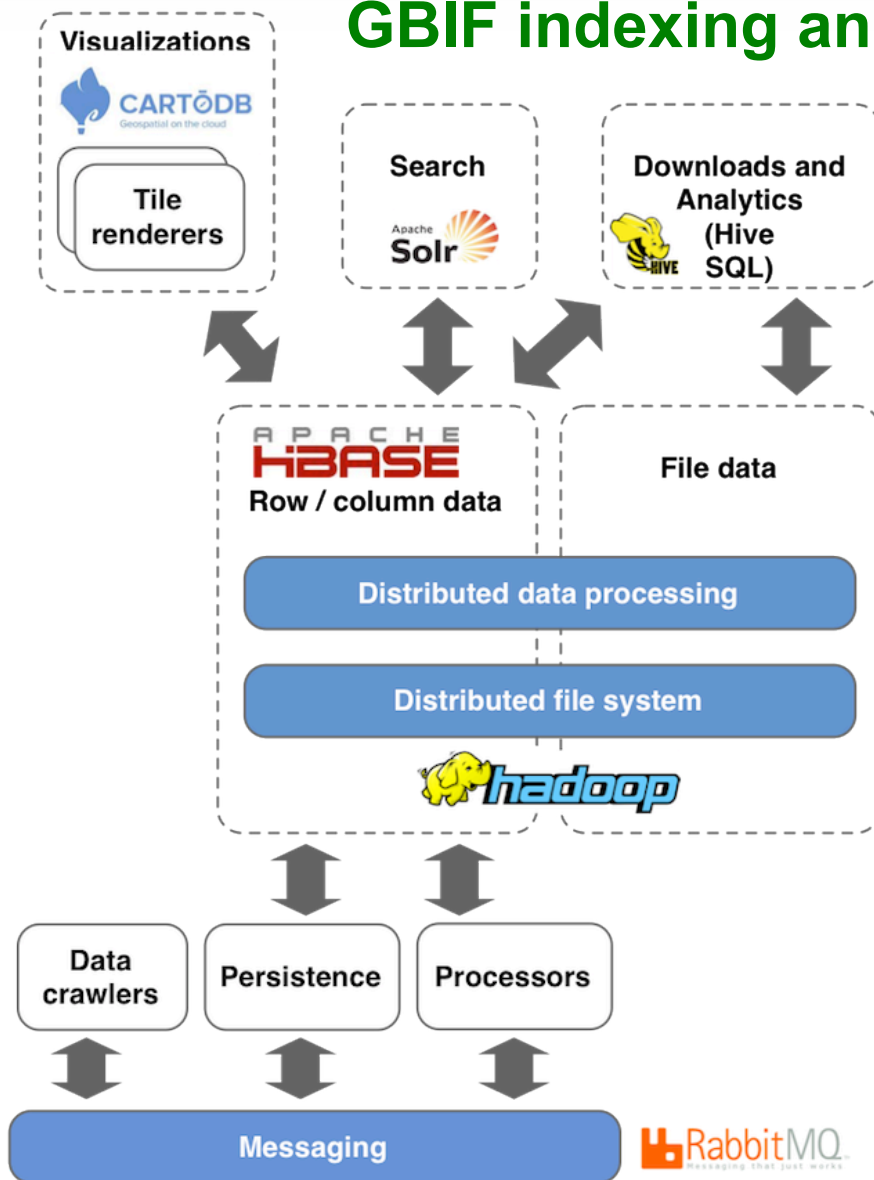
Central data portal

Central data index

Distributed data publishers



# GBIF indexing and portal ICT



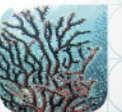
## The GBIF data standards (primary data)

### Access to Biological Collection Data (ABCD)

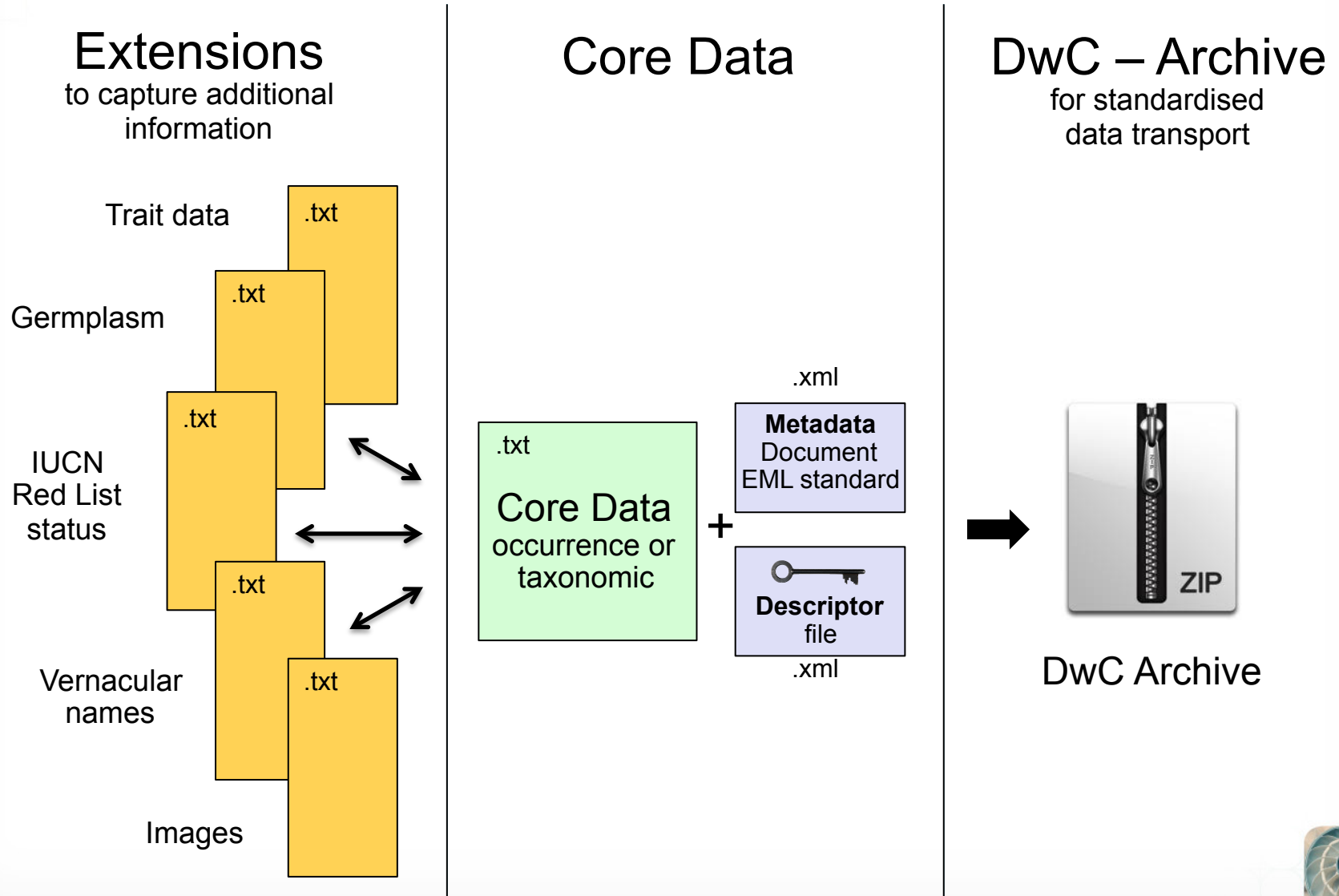
- Full coverage approach: ABCD is comprehensive and therefore complex. It explicitly aims to define the semantics of all elements.
- Element groups: nearly 1200 concepts provided
- Metadata embedded

### Darwin Core (DwC)

- Body of data standards which function as an extension of Dublin Core for biodiversity informatics
- 9 categories, about 180 terms
- Additional info in data extensions



# The GBIF data format (DarwinCore-Archives)



# Example of the core data in GBIF



## Occurrence

CATALOGNUMBER	NMR996800020933
DISPOSITION	in collection
INDIVIDUALCOUNT	1
PREPARATIONS	whole animal

## Event

DAY	01
EVENTDATE	1971-05-01
MONTH	05
YEAR	1971

## Location

CONTINENT	Europe
COUNTRY	Netherlands
COUNTRYCODE	NL-NB
DECIMALLATITUDE	51.630592
DECIMALLONGITUDE	5.135164

GEOETICDATUM	WGS84
GEOREFERENCEDBY	F.J.A. Sliker
LOCALITY	Udenhout, De Brand
MINIMUMELEVATIONINMETERS	10
STATEPROVINCE	Noord-Brabant

## Taxon

CLASS	Insecta
FAMILY	Pieridae
GENUS	Anthocharis
KINGDOM	Animalia
NOMENCLATURALCODE	ICZN
ORDER	Lepidoptera
ORIGINALNAMEUSAGE	Oranjetipje
PHYLUM	Arthropoda
SCIENTIFICNAME	<i>Anthocharis cardamines</i> (Linnaeus, 1758)
SCIENTIFICNAMEAUTHORSHIP	(Linnaeus, 1758)
SPECIFICEPTHET	cardamines



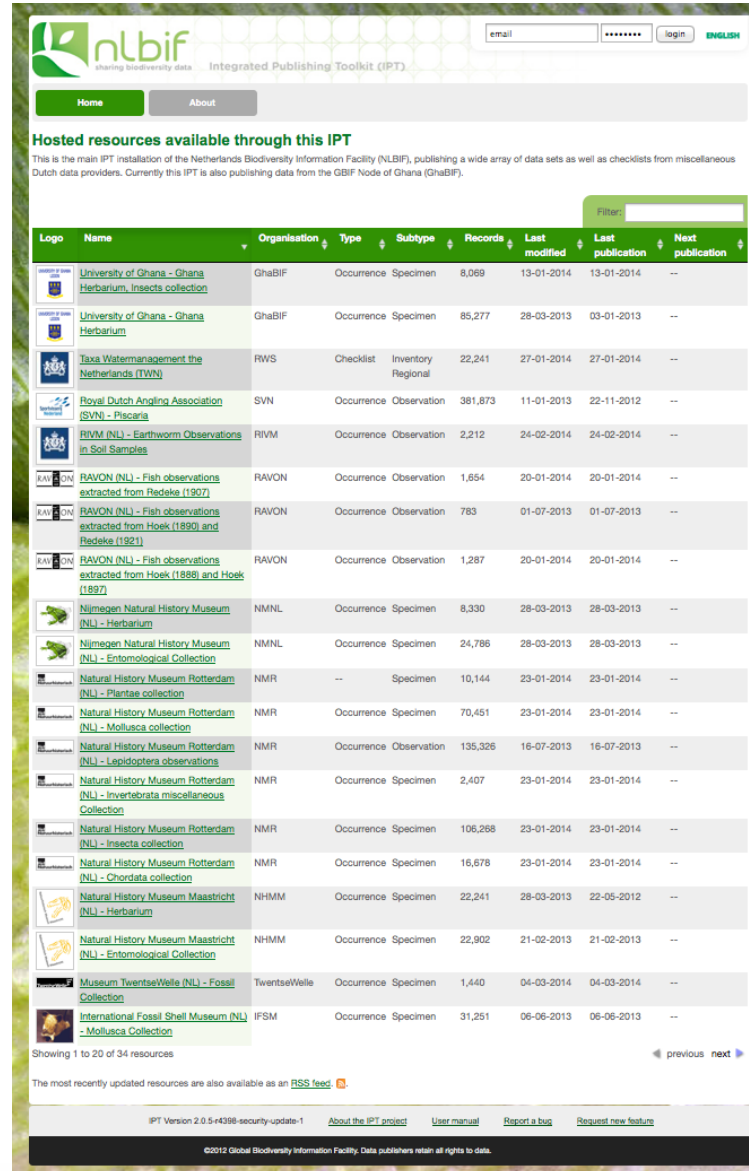
# Data publishing in GBIF

A distributed network of “Integrated Data Publishing Toolkits”





















Simple, easy to deploy

Primary data + Metadata

Stable url for your DarwinCore Archive



The screenshot shows the nlbif IPT interface. At the top, there is a search bar and a language selector set to 'ENGLISH'. Below the navigation tabs, a section titled 'Hosted resources available through this IPT' provides a brief description. The main content is a table listing various data resources with columns for Logo, Name, Organisation, Type, Subtype, Records, Last modified, Last publication, and Next publication. The table lists 34 resources, including collections from the University of Ghana, RWS, SVN, RIVM, RAVON, and various Dutch museums like Nijmegen, Rotterdam, and Maastricht.

Logo	Name	Organisation	Type	Subtype	Records	Last modified	Last publication	Next publication
	<a href="#">University of Ghana - Ghana Herbarium, Insects collection</a>	GhaBIF	Occurrence	Specimen	8,069	13-01-2014	13-01-2014	--
	<a href="#">University of Ghana - Ghana Herbarium</a>	GhaBIF	Occurrence	Specimen	85,277	28-03-2013	03-01-2013	--
	<a href="#">Taxa Watermanagement the Netherlands (TWN)</a>	RWS	Checklist	Inventory Regional	22,241	27-01-2014	27-01-2014	--
	<a href="#">Royal Dutch Aradina Association (SVN) - Piscaria</a>	SVN	Occurrence	Observation	381,873	11-01-2013	22-11-2012	--
	<a href="#">RIVM (NL) - Earthworm Observations in Soil Samples</a>	RIVM	Occurrence	Observation	2,212	24-02-2014	24-02-2014	--
	<a href="#">RAVON (NL) - Fish observations extracted from Redeke (1907)</a>	RAVON	Occurrence	Observation	1,654	20-01-2014	20-01-2014	--
	<a href="#">RAVON (NL) - Fish observations extracted from Hoek (1890) and Redeke (1921)</a>	RAVON	Occurrence	Observation	783	01-07-2013	01-07-2013	--
	<a href="#">RAVON (NL) - Fish observations extracted from Hoek (1888) and Hoek (1897)</a>	RAVON	Occurrence	Observation	1,287	20-01-2014	20-01-2014	--
	<a href="#">Nijmegen Natural History Museum (NL) - Herbarium</a>	NMNL	Occurrence	Specimen	8,330	28-03-2013	28-03-2013	--
	<a href="#">Nijmegen Natural History Museum (NL) - Entomological Collection</a>	NMNL	Occurrence	Specimen	24,786	28-03-2013	28-03-2013	--
	<a href="#">Natural History Museum Rotterdam (NL) - Plantae collection</a>	NMR	--	Specimen	10,144	23-01-2014	23-01-2014	--
	<a href="#">Natural History Museum Rotterdam (NL) - Mollusca collection</a>	NMR	Occurrence	Specimen	70,451	23-01-2014	23-01-2014	--
	<a href="#">Natural History Museum Rotterdam (NL) - Lepidoptera observations</a>	NMR	Occurrence	Observation	135,326	16-07-2013	16-07-2013	--
	<a href="#">Natural History Museum Rotterdam (NL) - Invertebrata miscellaneous Collection</a>	NMR	Occurrence	Specimen	2,407	23-01-2014	23-01-2014	--
	<a href="#">Natural History Museum Rotterdam (NL) - Insecta collection</a>	NMR	Occurrence	Specimen	106,268	23-01-2014	23-01-2014	--
	<a href="#">Natural History Museum Rotterdam (NL) - Chordata collection</a>	NMR	Occurrence	Specimen	16,678	23-01-2014	23-01-2014	--
	<a href="#">Natural History Museum Maastricht (NL) - Herbarium</a>	NHMM	Occurrence	Specimen	22,241	28-03-2013	22-05-2012	--
	<a href="#">Natural History Museum Maastricht (NL) - Entomological Collection</a>	NHMM	Occurrence	Specimen	22,902	21-02-2013	21-02-2013	--
	<a href="#">Museum TwenteWelle (NL) - Fossil Collection</a>	TwenteWelle	Occurrence	Specimen	1,440	04-03-2014	04-03-2014	--
	<a href="#">International Fossil Shell Museum (NL) - Mollusca Collection</a>	IFSM	Occurrence	Specimen	31,251	06-06-2013	06-06-2013	--

Showing 1 to 20 of 34 resources ◀ previous next ▶

The most recently updated resources are also available as an [RSS feed](#).

IPT Version 2.0.5-4398-security-update-1 [About the IPT project](#) [User manual](#) [Report a bug](#) [Request new feature](#)

©2012 Global Biodiversity Information Facility. Data publishers retain all rights to data.





# The Tool!

Integrated  
Publishing  
Toolkit

See examples at:

[ipt.nlbif.nl](http://ipt.nlbif.nl)

<http://www.cgn.ipt.wur.nl/>

<http://ipt.ibed.uva.nl>

<http://ipt.ibed.uva.nl>



Logged in as c.h.j.hof@uva.nl [Account](#) [Logout](#)

nlbif sharing biodiversity data Integrated Publishing Toolkit (IPT)

Home **Manage Resources** Administration About

**Geographic Coverage: *University of Amsterdam, Páramo vegetation research, Western Andean Cordillera***

Drag the markers or fill in the fields to set the geographic bounding box of the area covered by the resource.

2000 km 2000 mi North Pacific Ocean South Atlantic Ocean

United States Mexico Venezuela Colombia Brazil Peru Bolivia Chile

Spain North Atlantic Ocean South Atlantic Ocean

Algeria Libya Egypt Mauritania Mali Niger Chad Sudan Nigeria Ethiopia DR Congo Kenya Tanzania Angola Namibia Botswana

Map

Google Map data ©2012 MapLink, Tele Atlas

Set global coverage

West  East

South  North

Description


**Section**

- [Basic Metadata](#)
- [Geographic Coverage](#)
- [Taxonomic Coverage](#)
- [Temporal Coverage](#)
- [Keywords](#)
- [Associated Parties](#)
- [Project Data](#)
- [Sampling Methods](#)
- [Citations](#)
- [Collection Data](#)
- [External links](#)
- [Additional Metadata](#)

**Metadata Authoring  
Primary Biodiversity Data  
Species Checklists**




# The Tool!


Logged in as [c.h.j.hof@uva.nl](#) [Account](#) [Logout](#)

Home
Manage Resources
Administration
About

## Dutch Foundation for Applied Water Research (STOWA) - Limnodata Neerlandica [Edit](#)

The Limnodata Neerlandica data set contains the data of more than 30 years systematic and project based sampling of Dutch, mainly freshwater, waterbodies. Data originate from 37 data providers; 26 water boards, the provinces and the Ministry of Transport, Public Works and Water Management. Over 20,000 sample points were used. The main purpose of the dataset is monitoring the quality of Dutch surface waters. Next to the biological characteristics also a-biotic data were sampled and stored alongside the biotic data. Observations are available from phytoplankton, diatoms, aquatic and riparian plants, macro-invertebrates and fish. The data were collected on a yearly base. The taxonomy has been validated according to the Dutch standardised checklist of aquatic species.

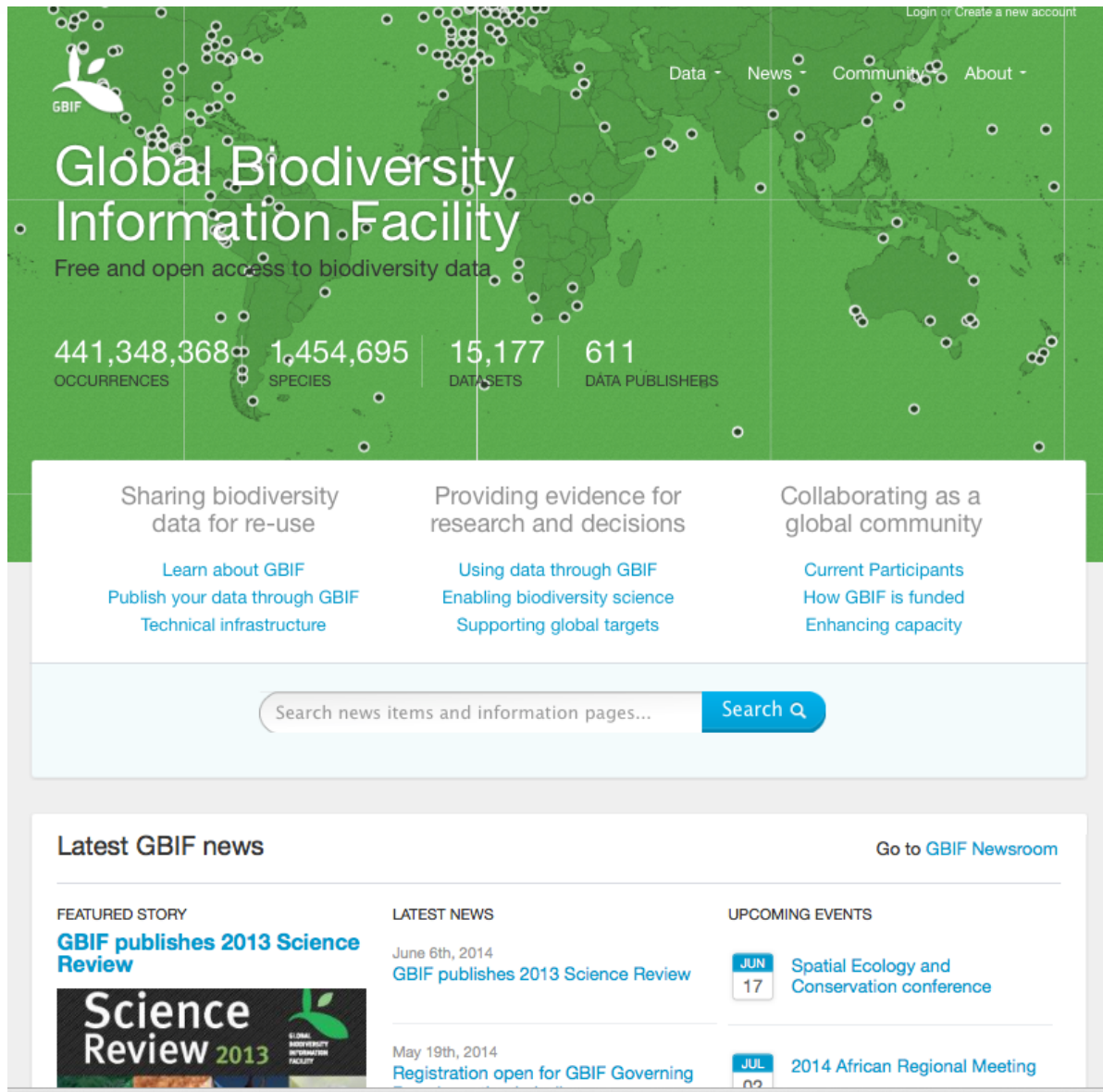


<b>Summary</b>	<p><b>Keywords</b> Limnofauna; aquatic; riparian; plants; phytoplankton; diatoms; zooplankton; macrofauna; fish; Netherlands; monitoring</p> <p><b>Metadata Language</b> English</p> <p><b>Resource Language</b> English</p> <p><b>Last Publication</b> Version 3 from Nov 5, 2012</p> <p><b>Darwin Core Archive</b> <a href="#">download</a> (69 MB) 2,942,518 records</p> <p><b>EML</b> <a href="#">download</a> (7 KB)</p> <p><b>RTF</b> <a href="#">download</a> (8 KB)</p> <p><b>GBIF Registration</b> <a href="#">37f48e00-1fe8-11dc-b461-b8a03c50a862</a></p> <p><b>Organisation</b> <a href="#">Dutch Foundation for Applied Water Research</a></p> <p><b>Endorsing Node</b></p>
<b>External Links</b>	<p><b>Resource Homepage</b> <a href="http://www.limnodata.nl/">http://www.limnodata.nl/</a></p>
<b>Resource Contact</b>	<p><b>Name</b> Bas van der Wal</p> <p><b>Organisation</b> STOWA</p> <p><b>Address</b> P.O. Box 2180, Amersfoort, NETHERLANDS, Postal Code: NL-3800 CD</p> <p><b>Contact</b> <a href="mailto:stowa@nlbif.nl">stowa@nlbif.nl</a></p> <p><b>Home Page</b> <a href="http://www.stowa.nl">http://www.stowa.nl</a></p>
<b>Resource Creator</b>	<p><b>Name</b> Roel Knoben</p> <p><b>Organisation</b> Royal Haskoning DHV</p> <p><b>Address</b> P.O. Box 525, Den Bosch, NETHERLANDS, Postal Code: NL-5201 AM</p> <p><b>Contact</b> <a href="mailto:stowa@nlbif.nl">stowa@nlbif.nl</a></p> <p><b>Home Page</b> <a href="http://www.stowa.nl">http://www.stowa.nl</a></p>



# The Portal

<http://www.gbif.org>



Global Biodiversity Information Facility

Free and open access to biodiversity data

441,348,368 OCCURRENCES | 1,454,695 SPECIES | 15,177 DATASETS | 611 DATA PUBLISHERS


Sharing biodiversity data for re-use  
[Learn about GBIF](#)  
[Publish your data through GBIF](#)  
[Technical infrastructure](#)

Providing evidence for research and decisions  
[Using data through GBIF](#)  
[Enabling biodiversity science](#)  
[Supporting global targets](#)

Collaborating as a global community  
[Current Participants](#)  
[How GBIF is funded](#)  
[Enhancing capacity](#)

Search news items and information pages... [Search](#)

Latest GBIF news [Go to GBIF Newsroom](#)


FEATURED STORY  
**GBIF publishes 2013 Science Review**  


LATEST NEWS  
 June 6th, 2014  
[GBIF publishes 2013 Science Review](#)

UPCOMING EVENTS  
 JUN 17 [Spatial Ecology and Conservation conference](#)  
 JUL 02 [2014 African Regional Meeting](#)



## The data in GBIF



GBIF.ORG  
Free and open access to biodiversity data

[Data](#) - [News](#) - [Community](#) - [About](#) -

# Dutch Foundation for Applied Water R...

Occurrence dataset published by Dutch Foundation for Applied Water Research

**2.942.518**  
Occurrences

[View occurrences](#)

[Information](#) [Stats](#) [Activity](#)

### Summary

FULL TITLE

Dutch Foundation for Applied Water Research (STOWA) - Limnodata Neerlandica

DESCRIPTION

The Limnodata Neerlandica data set contains the data of more than 30 years systematic and project based sampling of Dutch, mainly freshwater, waterbodies. Data originate from 37 data providers; 26 water boards, the provinces and the Ministry of Transport, Public Works and Water Management. Over 20,000 sample points were used. The main purpose of the dataset is monitoring the quality of Dutch surface waters. Next to the biological characteristics also a-biotic data were sampled and stored alongside the biotic data. Observations are available from phytoplankton, diatoms, aquatic and riparian plants, macro-invertebrates and fish. The data were collected on a yearly base. The taxonomy has been validated according to the Dutch standardised checklist of aquatic species.

PURPOSE

Water management and research.

TEMPORAL COVERAGES

Date range: 1-jan-1980 - 31-dec-2010

LANGUAGE OF DATA

ADMINISTRATIVE CONTACT


[Bas van der Wal](#)

METADATA AUTHOR

[Bas van der Wal](#)

ORIGINATOR

[Roel Knobon](#)



PUBLISHED BY

[Dutch Foundation for Applied Water Research](#)

PUBLICATION DATE

5-nov-2012

REGISTRATION DATE

4-mei-2010

HOSTED BY

[Netherlands Biodiversity Information Facility \(NLBIF\)](#)

SERVED BY

[NLBIF IPT](#)

LINKS

[Dataset homepage](#)

ALTERNATIVE IDENTIFIERS

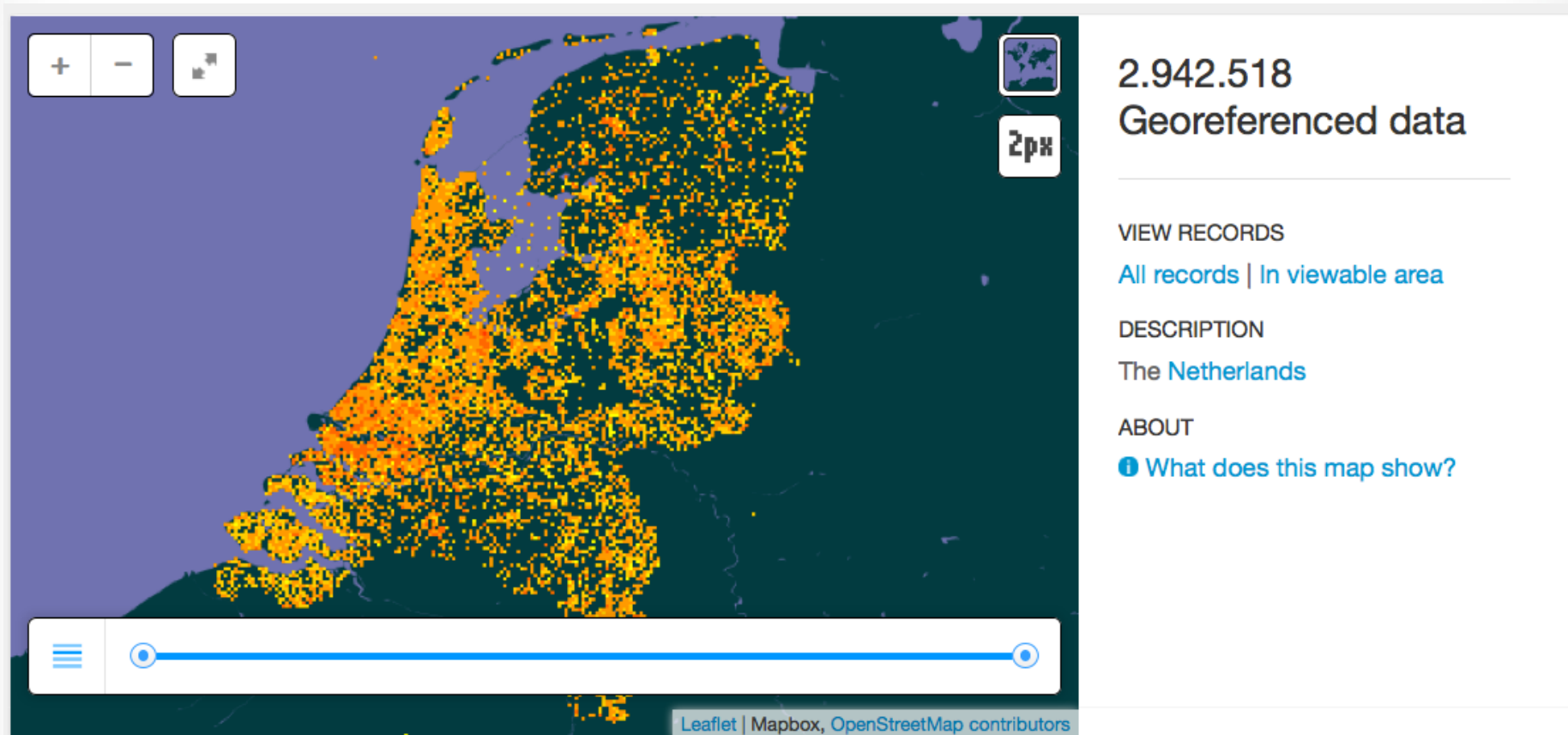
[GBIF Portal ID](#)  
<http://data.gbif.org/datasets/...>

EXTERNAL DATA

[Darwin Core Archive](#)



## The data in GBIF



# The data in GBIF



## Pinnularia gibba Ehrenberg

Species in GBIF Backbone Taxonomy

Plantae · Bacillariophyta · Bacillariophyceae · Naviculales · Pinnulariaceae · Pinnularia

612  
Occurrences

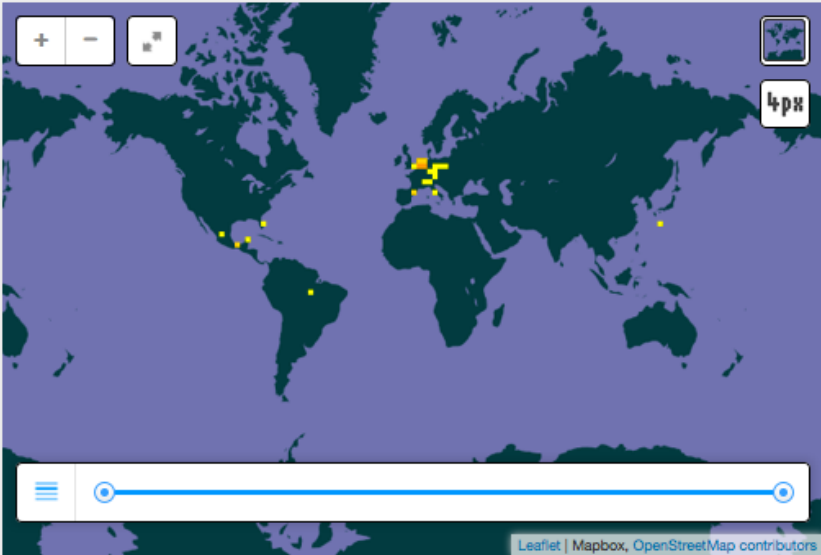
8  
Intraspecies

[View occurrences](#)

Information

### Overview

<p><b>FULL NAME</b></p> <p>Pinnularia gibba Ehrenberg</p>	<p><b>TAXONOMIC STATUS</b></p> <p>Accepted species</p> <p><b>ACCORDING TO</b></p> <p>The Catalogue of Life, 3rd January 2011</p>	<p><b>EXTERNAL LINKS</b></p> <ul style="list-style-type: none"> <li>· <a href="#">Encyclopedia of Life</a></li> <li>· <a href="#">Catalogue of Life</a></li> <li>· <a href="#">Biodiversity Heritage Library</a></li> </ul> <p><b>GBIF ID</b>    5268655</p>
---	--	--



Leaflet | Mapbox, OpenStreetMap contributors

### Georeferenced data

**VIEW RECORDS**

[All 577](#) | [In viewable area](#)

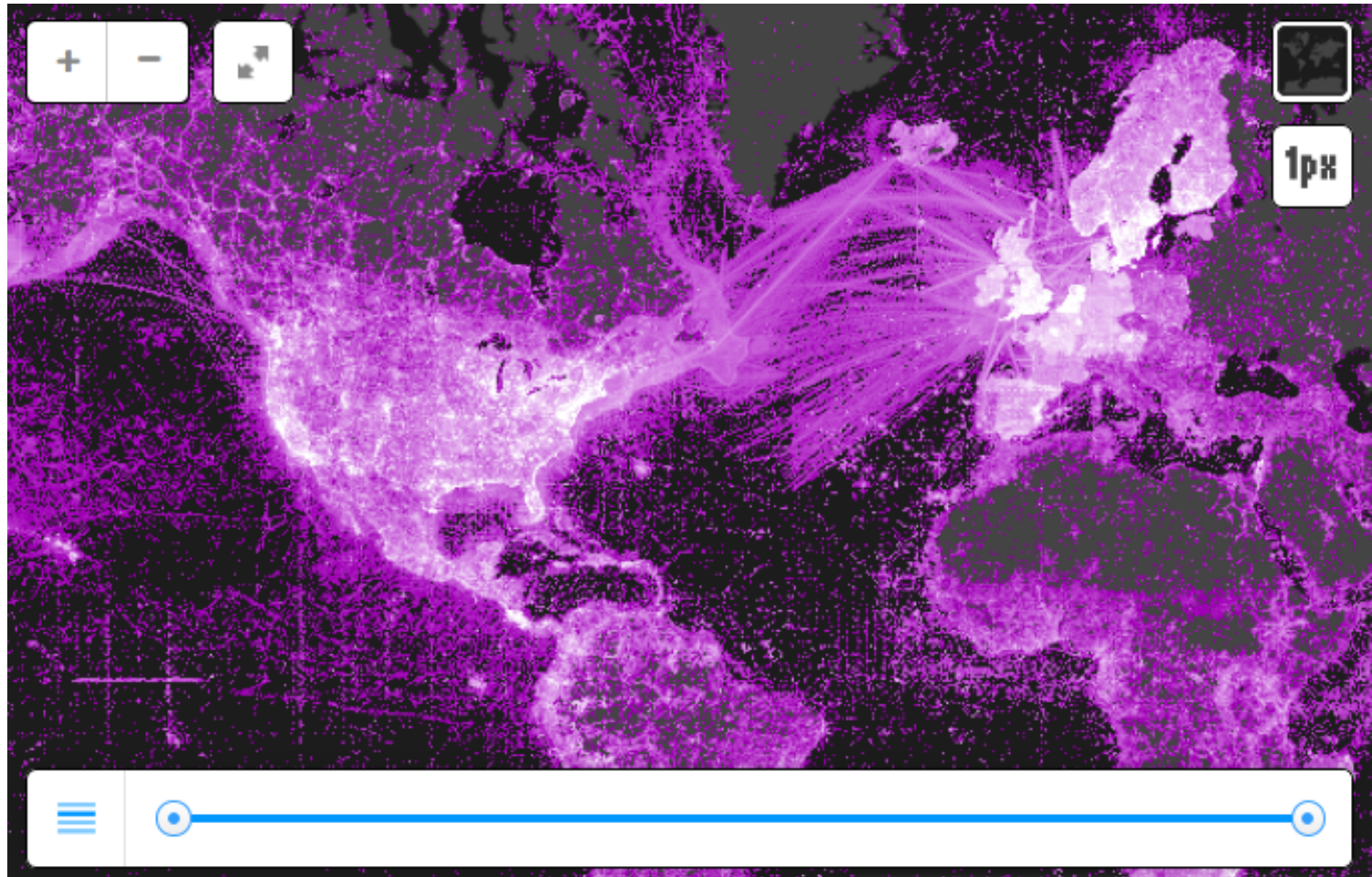
**DISTRIBUTIONS**

Text based [distributions](#) present in some sources.

Subordinate taxa

Classification





nature  
climate change

Home | Opinion & Analysis | Research | Current Issue | Archive

Archive > 2013 > July > Letters > Abstract

**ARTICLE PREVIEW**  
view full access options >


NATURE CLIMATE CHANGE | LETTER


## Quantifying the benefit of early mitigation in avoiding biodiversity loss

R. Warren, J. VanDerWal, J. Price, J. A. Welbergen, I. Atkinson, A. Osborn, A. Jarvis, L. P. Shoo, S. E. Williams & J. Lowe

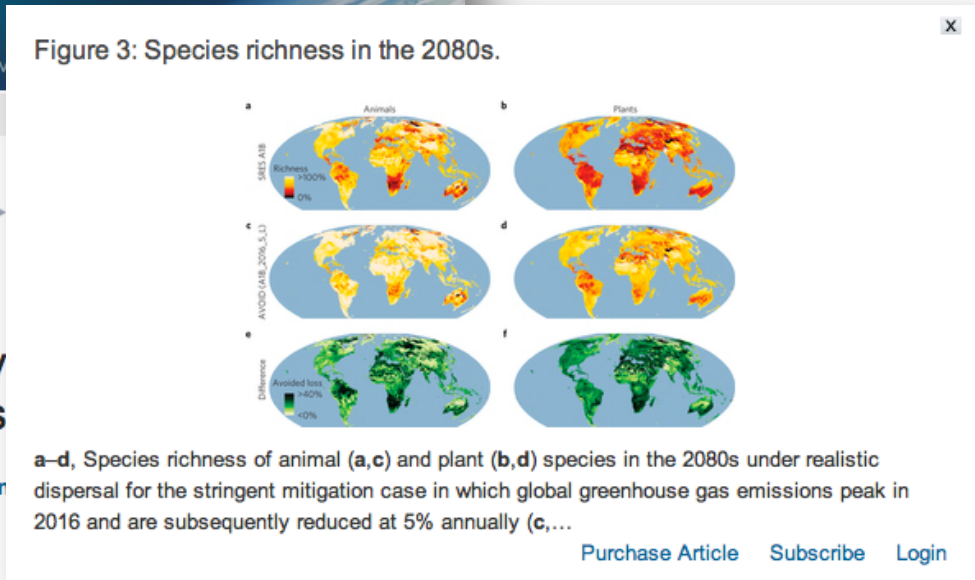
[Affiliations](#) | [Contributions](#) | [Corresponding author](#)

*Nature Climate Change* **3**, 678–682 (2013) | doi:10.1038/nclimate1887  
Received 09 August 2012 | Accepted 28 March 2013 | Published online 12 May 2013

 Citation

 Reprints

Climate change is expected to reduce biodiversity at global levels, including species-level extinctions of endemic species<sup>1, 2, 3, 4, 5, 6</sup>. However, if greenhouse gas emissions could reduce biodiversity loss, this could reduce biodiversity loss.



The information on the current location of common species of plants, mammals, birds, reptiles and amphibians was taken from around 170 million individual data records published freely online through GBIF by some 200 different institutions around the world. The records include museum specimens, data from scientific expeditions and the observations of thousands of volunteer 'citizen scientists'.





## The two big questions for GBIF:

To what extent do the data represent the true biodiversity of the world?

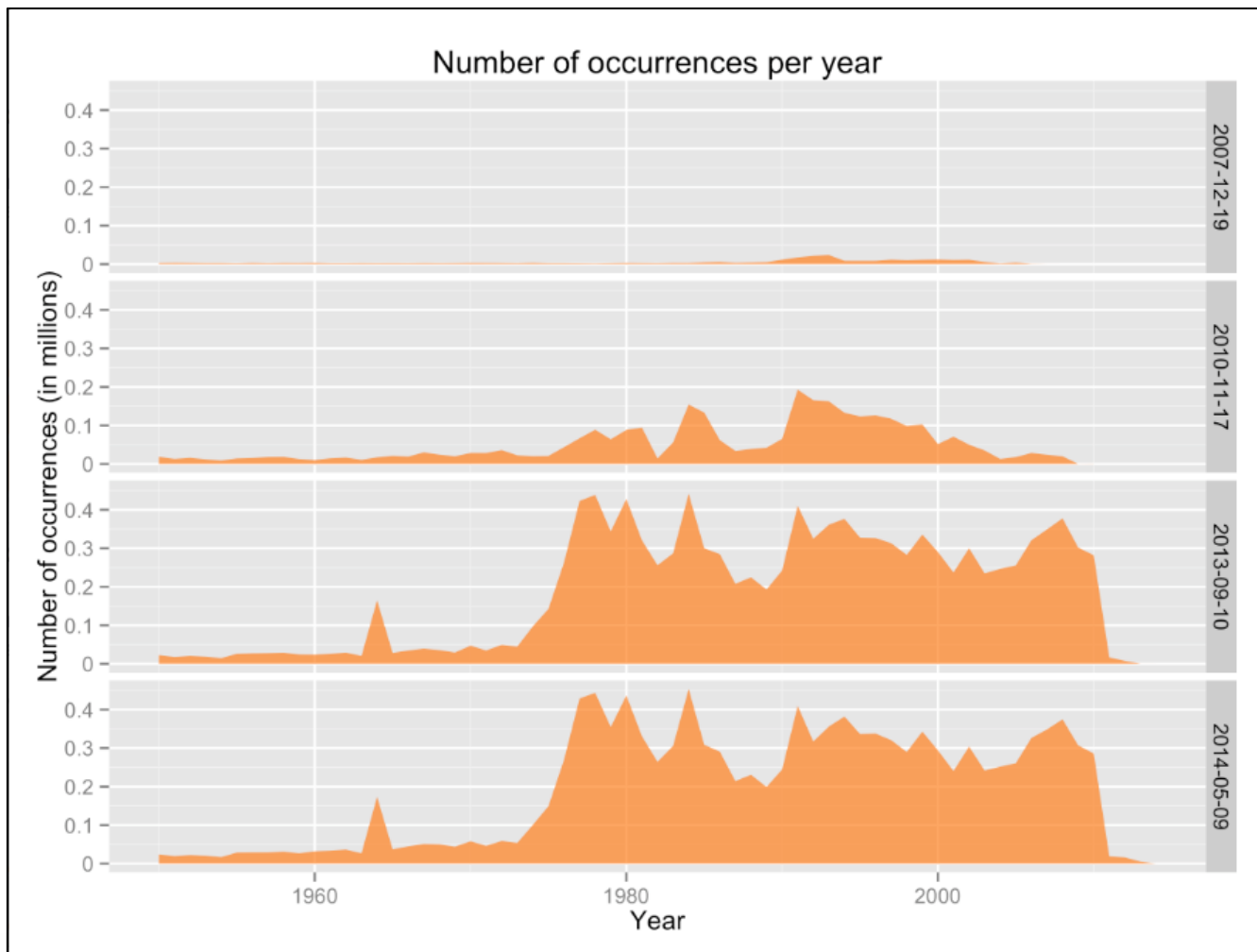
- Biases, gaps, disparities, selections, artifacts, etc.etc.
- How to determine the “fit for purpose”

What do the data tell us when combined with other data resources?

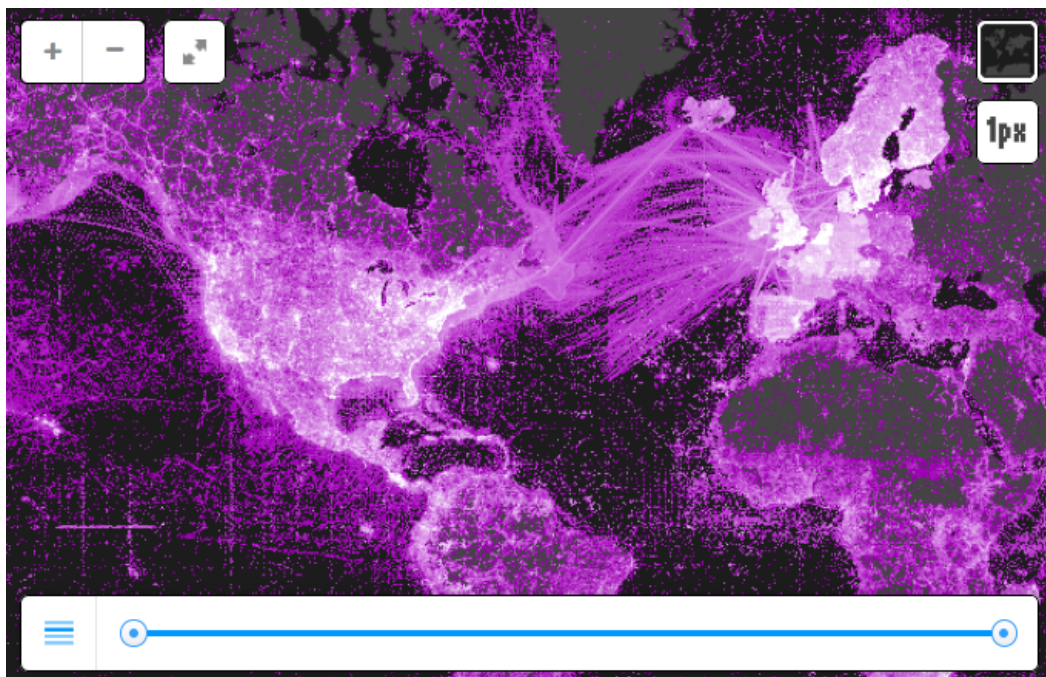
- GIS, genetics, habitats, nature conservation, etc. etc.
- Remote sensing, infrastructure, urbanisation, crops, etc. etc.



For example; Occurrence records by year of occurrence for the Netherlands



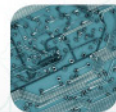
# Challenges...



+ 400 million open access biodiversity records in GBIF

Data are being used in scientific publications, but what else is in there?

- Advance data statistics...
- Combining with other info domains (geography, climate, infrastructure, etc.)
- Data streamlining...
- “Fit for purpose” how to meet the needs of data users...



<http://www.gbif.org>

