

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 biodiversity?



"Biological diversity" means the variability among living organisms from all sources including, inter alia, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems....





Biodiversity data are....



Natural history collections



Living collections





Biodiversity data are....



Systematic monitoring + research data



Citizen science data





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 of GBIF



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

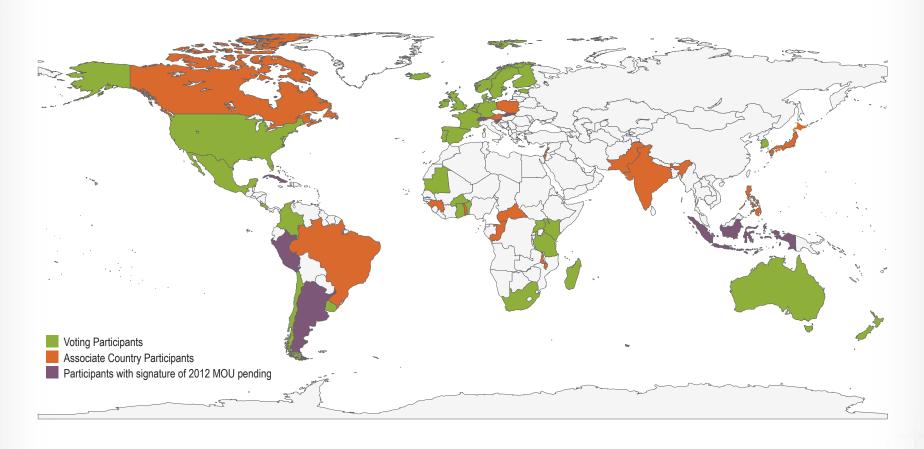
2012

Currently structurally financed by OCW, hosted by UvA





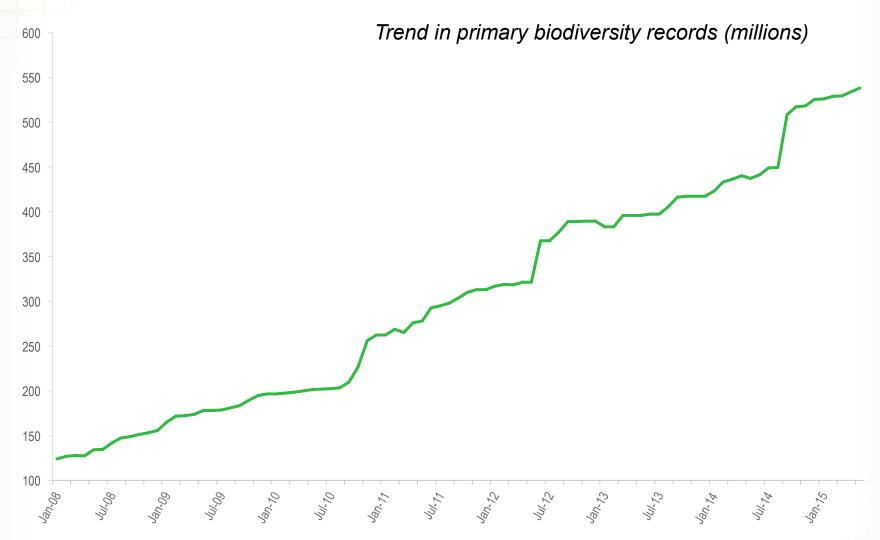
GBIF country participants







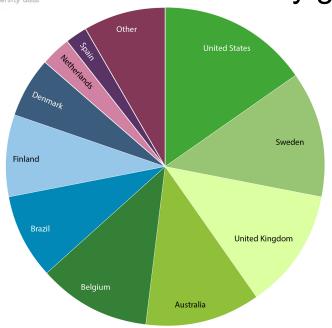
Data published through GBIF.org





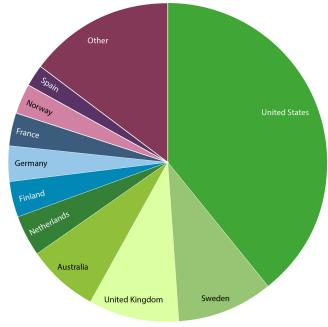


Data by gbif participant



Number of new records published—Top 10 participant Countries (1 Jan to 31 May 2015)

1. United States	2,951,100	6. Brazil	1,660,476
2. Sweden	2,476,199	7. Finland	1,572,265
3. United Kingdom	2,372,307	8. Denmark	1,187,322
4. Australia	2,260,641	9. Netherlands	630,252
5. Belgium	2,230,015	10. Spain	391,161



Total number of records published—Top 10 Participant Countries (as of 31 May 2015)

	,	• ,	
1. United States	209,679,618	6. Finland	20,036,123
2. Sweden	51,325,126	7. Germany	18,909,265
3. United Kingdom	49,543,695	8. France	17,577,672
4. Australia	38,955,381	9. Norway	17,425,011
5. Netherlands	21,578,737	10. Spain	10,578,485





The primary GBIF data:

What is a record in GBIF?

GBIF hosts so called "occurrence data"

 \Rightarrow

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

Taxonomic information



Geographic information



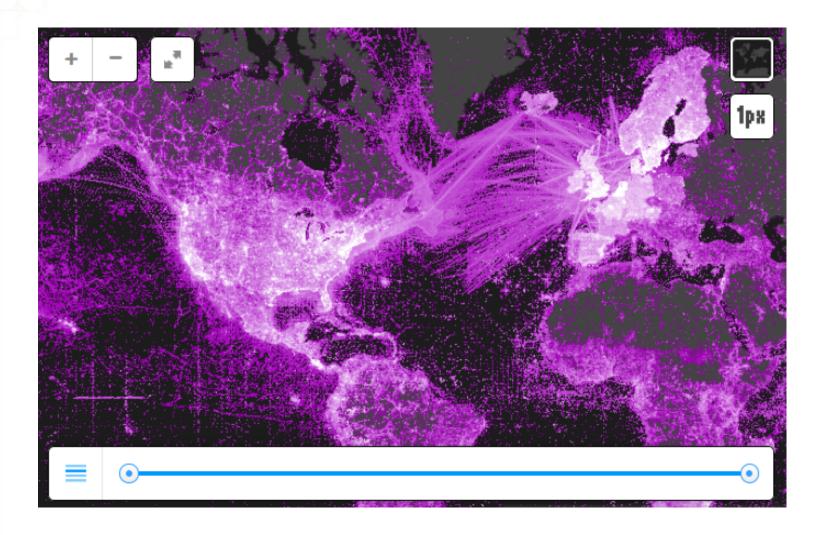
Temporal information

01 May 1971 00:00:00





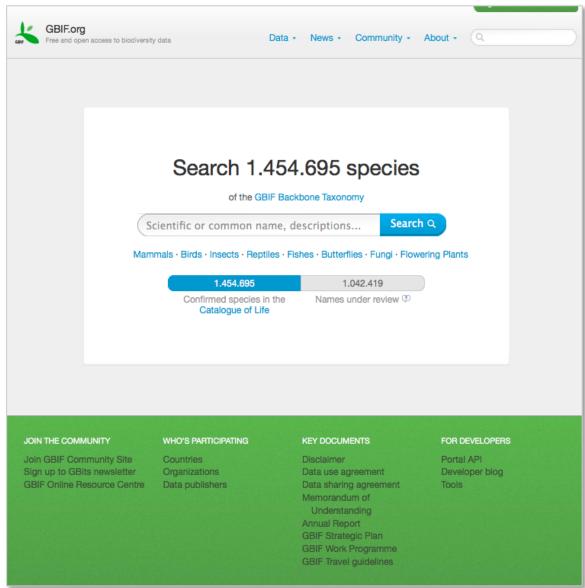
http://www.gbif.org/occurrence







http://www.gbif.org/species





Asellus aquaticus (Linnaeus, 1758)

Species in GBIF Backbone Taxonomy

Animalia · Arthropoda · Malacostraca · Isopoda · Asellidae · Asellus

View occurrences

12

45.358

Information

Search for:



Overview

FULL NAME

Asellus aquaticus (Linnaeus, 1758)

COMMON NAMES

- zoetwaterpissebed nid
- o?liczka pospolita pol
- ośliczka pospolita pol

more »

SYNONYMS

· Oniscus aquaticus Linnaeus, 1758

TAXONOMIC STATUS

Accepted species

ACCORDING TO

The Catalogue of Life, 3rd January 2011

PUBLISHED IN

Linnaeus, C. (1758). Systema Naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Editio decima, reformata. Laurentius Salvius: Holmiae. ii, 824 pp.

HABITAT

Not terrestrial, Not marine

EXTERNAL LINKS

- Encyclopedia of Life
- Catalogue of Life
- Biodiversity Heritage Library

GBIF ID 2206247



Georeferenced data

VIEW RECORDS

All 44.432 | In viewable area

DISTRIBUTIONS

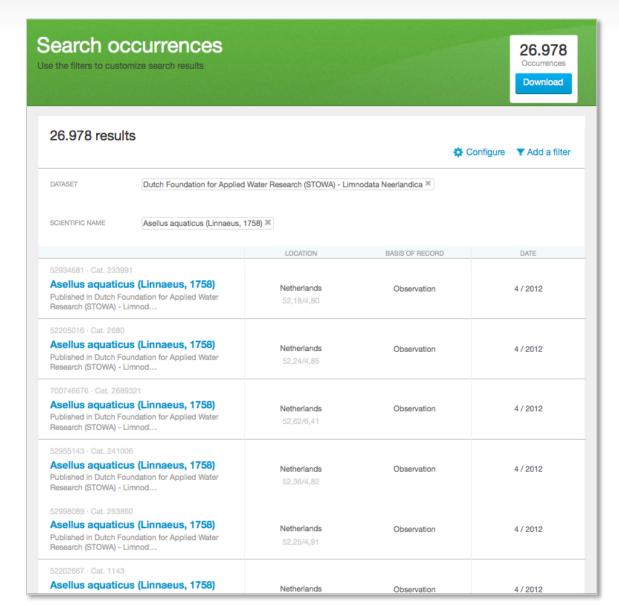
Text based distributions present in some sources.





Find....
Filter
&
Download

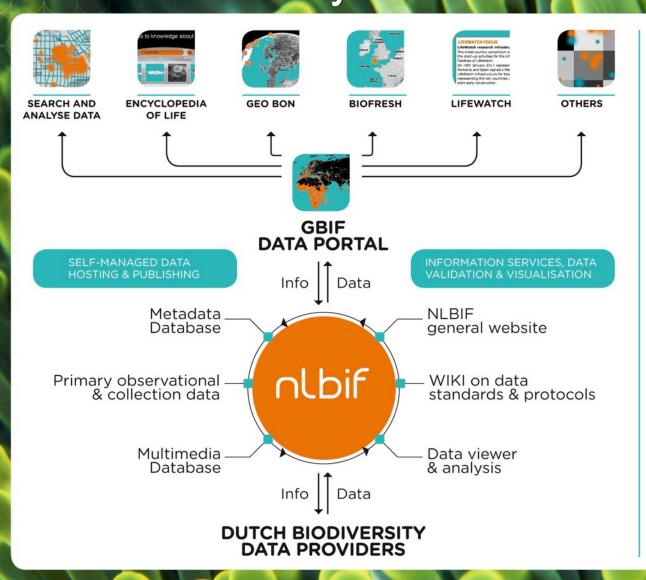








Netherlands Biodiverity Information Facility



GBIF data flow:

presentation

indexing

publishing

registration (+ archiving)

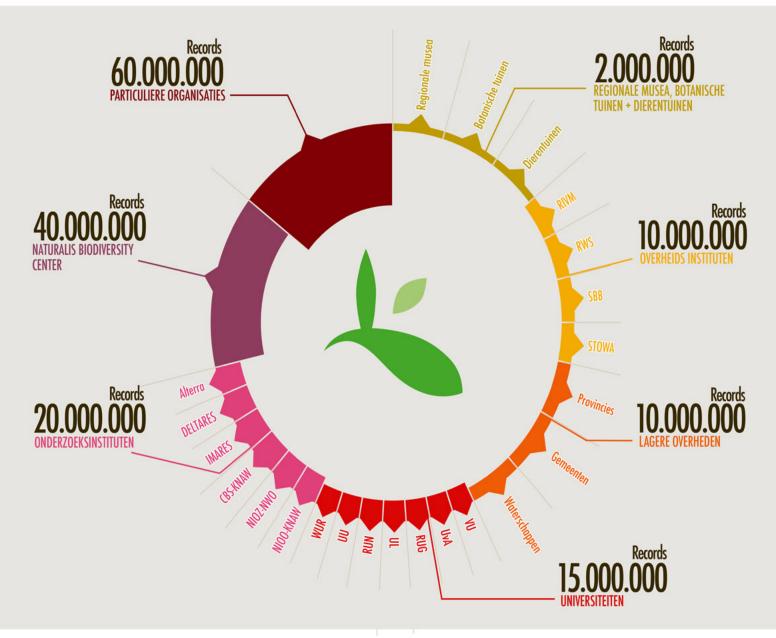
standardisation

data generation

NLBIF - DATA LANDSCAPE

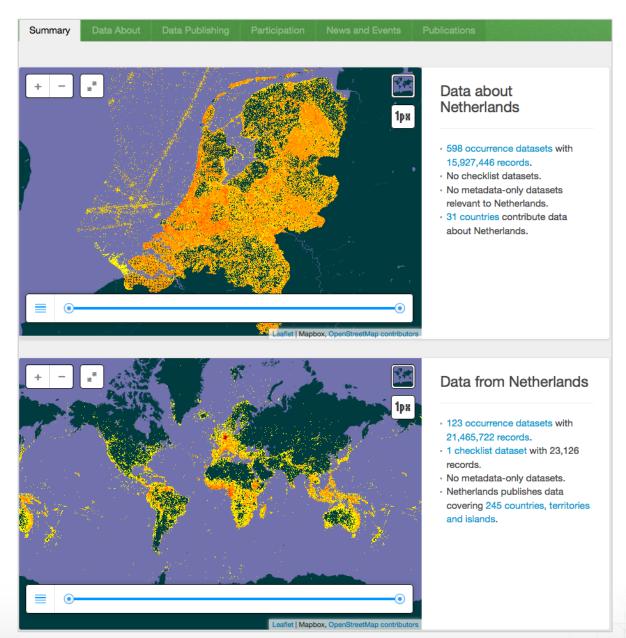


LET OP: VEEL GEGEVENS ZUN BU BENADERING





http://www.gbif.org/country/NL/summary





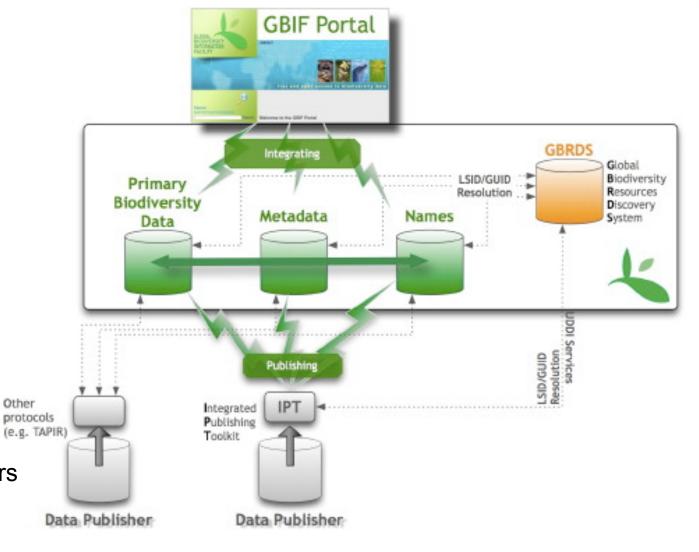


The GBIF infrastructure

Central data portal

Central data index

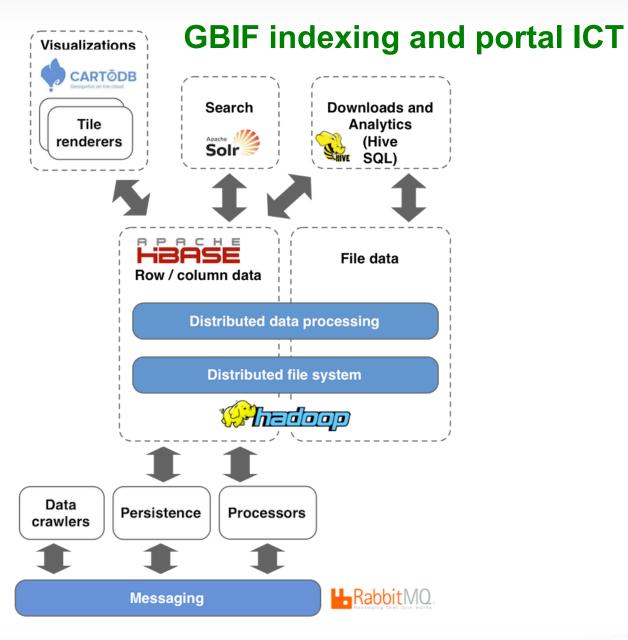
Distributed data publishers





Other







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)

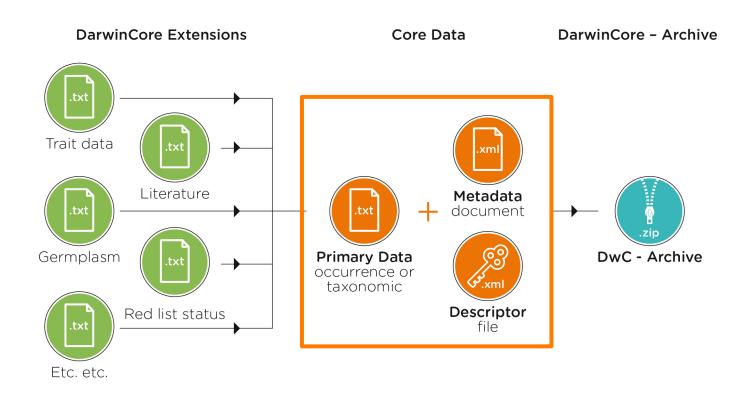
- Body of data standards which function as an extension of Dublin Core for biodiversity informatics
- 9 categories, about 180 terms
- Additional info in data extensions





The GBIF data axchange format (DarwinCore-Archives)

The generation and components of GBIF DarwinCore - Archives (DwC-A)







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	

GEODETICDATUM WGS84 GEOREFERENCEDBY F.J.A. Slieker 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 Anthocharis cardamines (Linnaeus, 1758) SCIENTIFICNAMEAUTHORSHIP (Linnaeus, 1758) SPECIFICEPITHET cardamines		
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 Anthocharis cardamines (Linnaeus, 1758) SCIENTIFICNAMEAUTHORSHIP (Linnaeus, 1758)	GEODETICDATUM	WGS84
MINIMUMELEVATIONINMETERS 10 STATEPROVINCE Noord-Brabant Taxon CLASS Insecta FAMILY Pieridae GENUS Anthocharis KINGDOM Animalia NOMENCLATURALCODE ICZN ORDER Lepidoptera ORIGINALNAMEUSAGE Oranjetipje PHYLUM Arthropoda SCIENTIFICNAME Anthocharis cardamines (Linnaeus, 1758) SCIENTIFICNAMEAUTHORSHIP (Linnaeus, 1758)	GEOREFERENCEDBY	F.J.A. Slieker
Taxon CLASS Insecta FAMILY Pieridae GENUS Anthocharis KINGDOM Animalia NOMENCLATURALCODE ICZN ORDER Lepidoptera ORIGINALNAMEUSAGE Oranjetipje PHYLUM Arthropoda SCIENTIFICNAME Anthocharis cardamines (Linnaeus, 1758) SCIENTIFICNAMEAUTHORSHIP (Linnaeus, 1758)	LOCALITY	Udenhout, De Brand
Taxon CLASS Insecta FAMILY Pieridae GENUS Anthocharis KINGDOM Animalia NOMENCLATURALCODE ICZN ORDER Lepidoptera ORIGINALNAMEUSAGE Oranjetipje PHYLUM Arthropoda SCIENTIFICNAME Anthocharis cardamines (Linnaeus, 1758) SCIENTIFICNAMEAUTHORSHIP (Linnaeus, 1758)	MINIMUMELEVATIONINMETERS	10
CLASS Insecta FAMILY Pieridae GENUS Anthocharis KINGDOM Animalia NOMENCLATURALCODE ICZN ORDER Lepidoptera ORIGINALNAMEUSAGE Oranjetipje PHYLUM Arthropoda SCIENTIFICNAME Anthocharis cardamines (Linnaeus, 1758) SCIENTIFICNAMEAUTHORSHIP (Linnaeus, 1758)	STATEPROVINCE	Noord-Brabant
CLASS Insecta FAMILY Pieridae GENUS Anthocharis KINGDOM Animalia NOMENCLATURALCODE ICZN ORDER Lepidoptera ORIGINALNAMEUSAGE Oranjetipje PHYLUM Arthropoda SCIENTIFICNAME Anthocharis cardamines (Linnaeus, 1758) SCIENTIFICNAMEAUTHORSHIP (Linnaeus, 1758)		
FAMILY Pieridae GENUS Anthocharis KINGDOM Animalia NOMENCLATURALCODE ICZN ORDER Lepidoptera ORIGINALNAMEUSAGE Oranjetipje PHYLUM Arthropoda SCIENTIFICNAME Anthocharis cardamines (Linnaeus, 1758) SCIENTIFICNAMEAUTHORSHIP (Linnaeus, 1758)	Taxon	
GENUS Anthocharis KINGDOM Animalia NOMENCLATURALCODE ICZN ORDER Lepidoptera ORIGINALNAMEUSAGE Oranjetipje PHYLUM Arthropoda SCIENTIFICNAME Anthocharis cardamines (Linnaeus, 1758) SCIENTIFICNAMEAUTHORSHIP (Linnaeus, 1758)	CLASS	Insecta
KINGDOM Animalia NOMENCLATURALCODE ICZN ORDER Lepidoptera ORIGINALNAMEUSAGE Oranjetipje PHYLUM Arthropoda SCIENTIFICNAME Anthocharis cardamines (Linnaeus, 1758) SCIENTIFICNAMEAUTHORSHIP (Linnaeus, 1758)	FAMILY	Pieridae
NOMENCLATURALCODE ICZN ORDER Lepidoptera ORIGINALNAMEUSAGE Oranjetipje PHYLUM Arthropoda SCIENTIFICNAME Anthocharis cardamines (Linnaeus, 1758) SCIENTIFICNAMEAUTHORSHIP (Linnaeus, 1758)	GENUS	Anthocharis
ORDER Lepidoptera ORIGINALNAMEUSAGE Oranjetipje PHYLUM Arthropoda SCIENTIFICNAME Anthocharis cardamines (Linnaeus, 1758) SCIENTIFICNAMEAUTHORSHIP (Linnaeus, 1758)	KINGDOM	Animalia
ORIGINALNAMEUSAGE Oranjetipje PHYLUM Arthropoda SCIENTIFICNAME Anthocharis cardamines (Linnaeus, 1758) SCIENTIFICNAMEAUTHORSHIP (Linnaeus, 1758)	NOMENCLATURALCODE	ICZN
PHYLUM Arthropoda SCIENTIFICNAME Anthocharis cardamines (Linnaeus, 1758) SCIENTIFICNAMEAUTHORSHIP (Linnaeus, 1758)	ORDER	Lepidoptera
SCIENTIFICNAME Anthocharis cardamines (Linnaeus, 1758) SCIENTIFICNAMEAUTHORSHIP (Linnaeus, 1758)	ORIGINALNAMEUSAGE	Oranjetipje
SCIENTIFICNAMEAUTHORSHIP (Linnaeus, 1758)	PHYLUM	Arthropoda
(Elinadas, 1700)	SCIENTIFICNAME	Anthocharis cardamines (Linnaeus, 1758)
SPECIFICEPITHET cardamines	SCIENTIFICNAMEAUTHORSHIP	(Linnaeus, 1758)
	SPECIFICEPITHET	cardamines



Integrated Publishing Toolkit

See examples at:

http://ipt.nlbif.nl

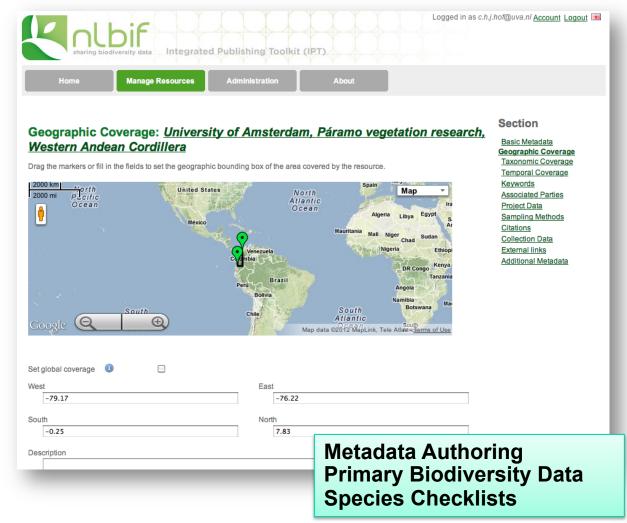
http://www.cgn.ipt.wur.nl/

http://ipt.ibed.uva.nl

http://ipt.nioz.nl

http://ipt.ibed.uva.nl

The data publishing tool!







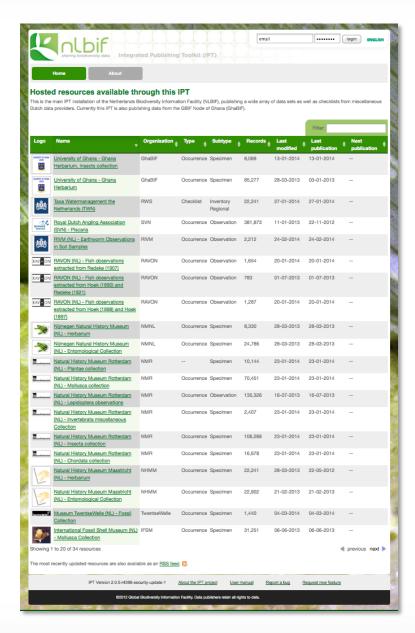
Data publishing in GBIF

A distributed network of "Integrated Data Publishing Toolkits"

Simple, easy to deploy

Primary data + Metadata

Stable url for your DarwinCore Archive

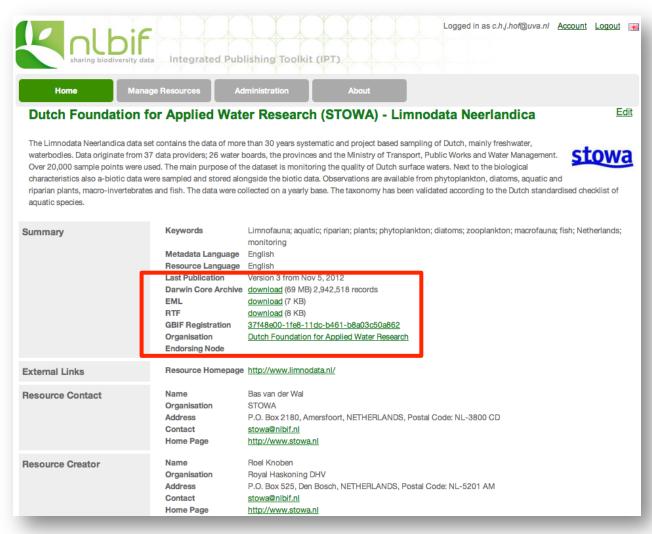






The Tool!

The data in the Toolkit of the dataprovider...

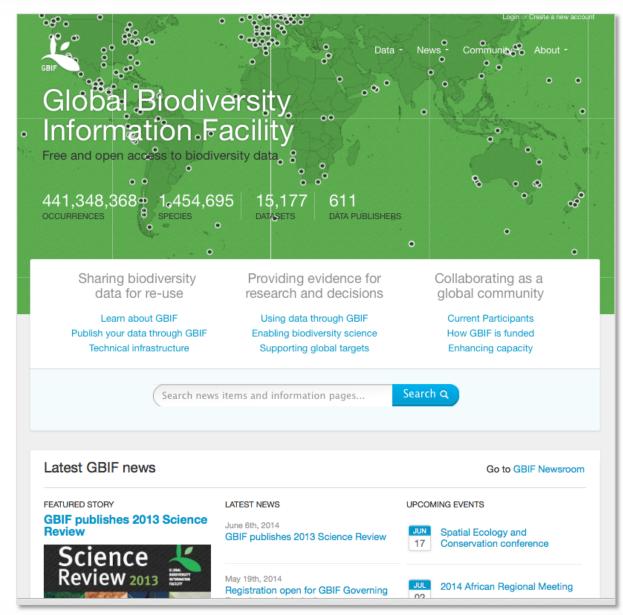






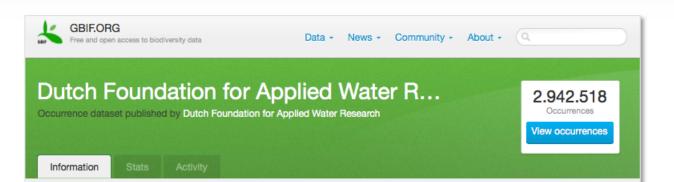
The Portal

http://www.gbif.org





The data in GBIF



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 Knoben

stowa

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

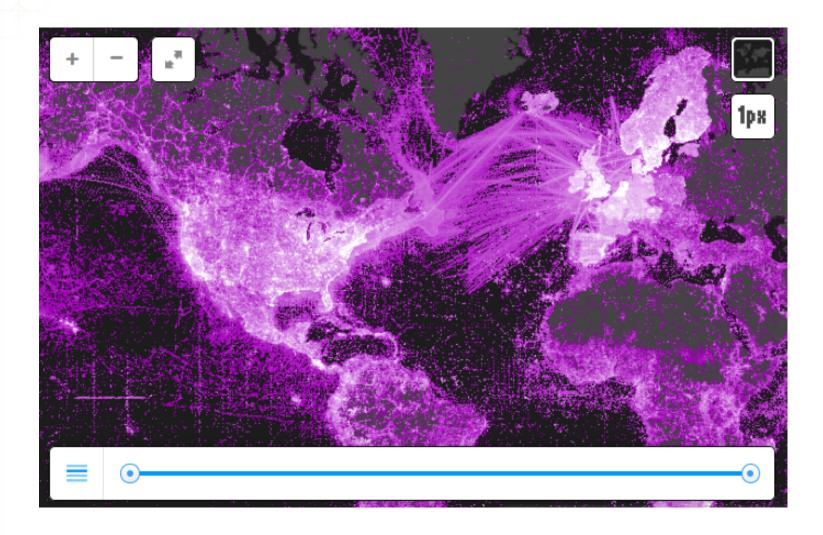




The data visualised in GBIF, etc.



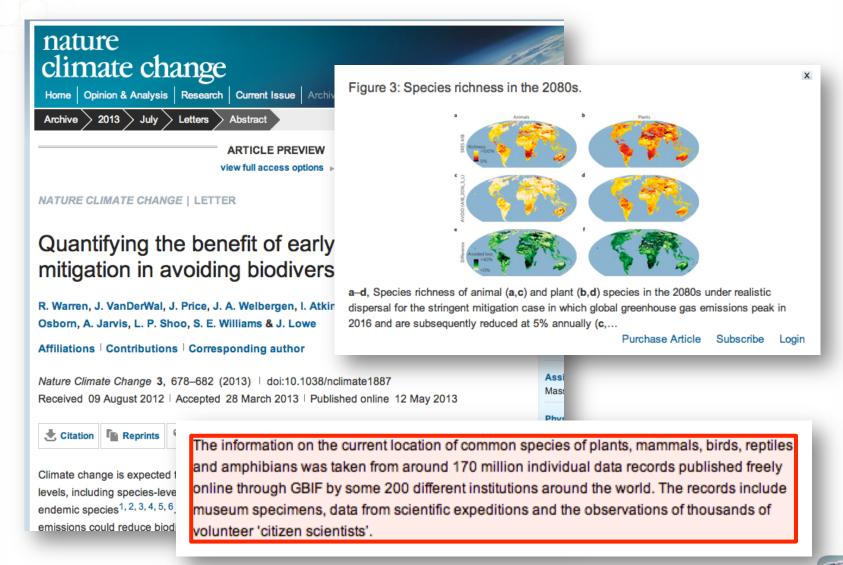






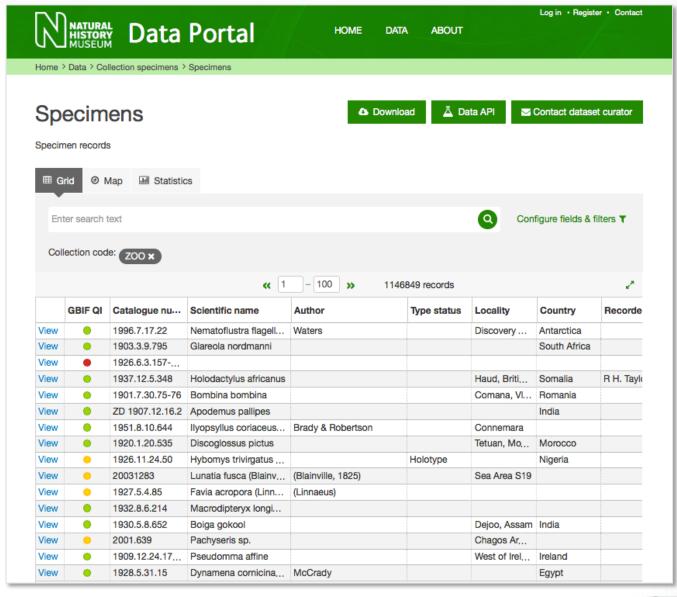


GBIF data use: Recent Papers





More and more (smart) use of the GBIF services







The two big questions for GBIF:

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

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

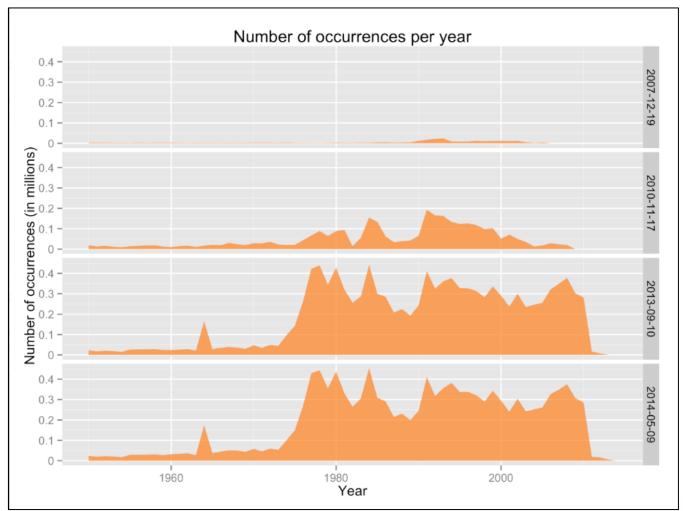
What do the data tell us when combined with other data resources?

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





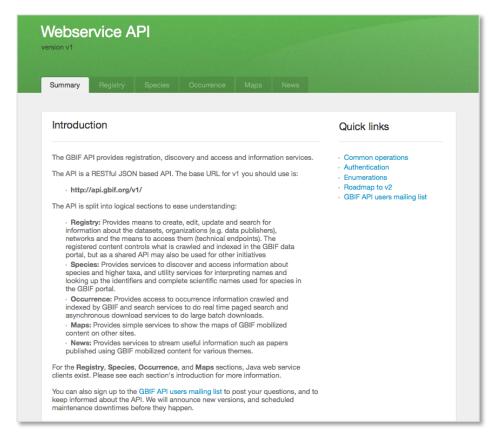
For example; Occurrence records by year of occurrence for the Netherlands







All GBIF data / metrics / metadata available (API):

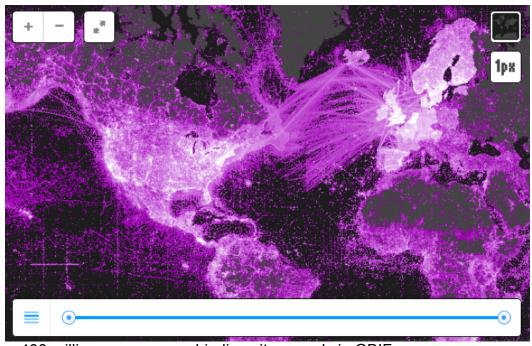


http://www.gbif.org/developer/summary





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

http://www.nlbif.nl

