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."
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 sustainable development.

Megascience Forum Working Group on Biological Informatics

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.

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

Voting Participants
Associate Country Participants
Participants with signature of 2012 MOU pending

Cees Hof
Netherlands Biodiversity Information Facility (NLBIF)

PIRE workshop 2015
Data published through GBIF.org

Trend in primary biodiversity records (millions)
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
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
5. Netherlands 21,578,737 10. Spain 10,578,485

Cees Hof
Netherlands Biodiversity Information Facility (NLBIF)
The primary GBIF 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**

![Butterfly](image1.png)

*Anthocharis cardamines (L., 1758) male*
*Netherlands, Noord-Brabant, Udenhout, De Brand. NMR 20933*

![Map](image2.png)

01 May 1971 00:00:00
http://www.gbif.org/occurrence
http://www.gbif.org/species
Search occurrences
Use the filters to customize search results

26.978 results

<table>
<thead>
<tr>
<th>DATASET</th>
<th>LOCATION</th>
<th>BASIS OF RECORD</th>
<th>DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asellus aquaticus (Linnaeus, 1758)</td>
<td>Netherlands</td>
<td>Observation</td>
<td>4/2012</td>
</tr>
<tr>
<td>Asellus aquaticus (Linnaeus, 1758)</td>
<td>Netherlands</td>
<td>Observation</td>
<td>4/2012</td>
</tr>
<tr>
<td>Asellus aquaticus (Linnaeus, 1758)</td>
<td>Netherlands</td>
<td>Observation</td>
<td>4/2012</td>
</tr>
<tr>
<td>Asellus aquaticus (Linnaeus, 1758)</td>
<td>Netherlands</td>
<td>Observation</td>
<td>4/2012</td>
</tr>
<tr>
<td>Asellus aquaticus (Linnaeus, 1758)</td>
<td>Netherlands</td>
<td>Observation</td>
<td>4/2012</td>
</tr>
</tbody>
</table>

Find.... Filter .... & Download
GBIF data flow:

- presentation
- indexing
- publishing
- registration (+ archiving)
- standardisation
- data generation
NLBIF - DATA LANDSCAPE

LET OP: VEEL GEGEVENS ZIJN BIJ BENADERING
(rough guess!)
http://www.gbif.org/country/NL/summary

**Data about Netherlands**
- 598 occurrence datasets with 15,927,446 records.
- No checklist datasets.
- No metadata-only datasets relevant to Netherlands.
- 31 countries contribute data about Netherlands.

**Data from Netherlands**
- 123 occurrence datasets with 21,465,722 records.
- 1 checklist dataset with 23,126 records.
- No metadata-only datasets.
- Netherlands publishes data covering 245 countries, territories and islands.
The GBIF infrastructure

Central data portal

Central data index

Distributed data publishers
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 exchange format (DarwinCore-Archives)

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

- DarwinCore Extensions
  - Trait data
  - Literature
  - Germplasm
  - Red list status
  - Etc. etc.

- Core Data
  - Primary Data occurrence or taxonomic
  - Metadata document
  - Descriptor file

- DarwinCore - Archive
  - DwC - Archive

Cees Hof
Netherlands Biodiversity Information Facility (NLBIF)
PIRE workshop 2015
Example of the core data in GBIF

### Occurrence

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CatalogNumber</td>
<td>NMR99560000020933</td>
</tr>
<tr>
<td>Disposition</td>
<td>in collection</td>
</tr>
<tr>
<td>IndividualCount</td>
<td>1</td>
</tr>
<tr>
<td>Preparations</td>
<td>whole animal</td>
</tr>
</tbody>
</table>

### Event

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day</td>
<td>01</td>
</tr>
<tr>
<td>EventDate</td>
<td>1971-05-01</td>
</tr>
<tr>
<td>Month</td>
<td>05</td>
</tr>
<tr>
<td>Year</td>
<td>1971</td>
</tr>
</tbody>
</table>

### Location

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continent</td>
<td>Europe</td>
</tr>
<tr>
<td>Country</td>
<td>Netherlands</td>
</tr>
<tr>
<td>CountryCode</td>
<td>NL-NB</td>
</tr>
<tr>
<td>DecimalLatitude</td>
<td>51.630592</td>
</tr>
<tr>
<td>DecimalLongitude</td>
<td>5.135164</td>
</tr>
</tbody>
</table>

### Taxon

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>Insecta</td>
</tr>
<tr>
<td>Family</td>
<td>Pieridae</td>
</tr>
<tr>
<td>Genus</td>
<td>Anthocaris</td>
</tr>
<tr>
<td>Kingdom</td>
<td>Animalia</td>
</tr>
<tr>
<td>NomenclaturalCode</td>
<td>IOZN</td>
</tr>
<tr>
<td>Order</td>
<td>Lepidoptera</td>
</tr>
<tr>
<td>OriginalNameUsage</td>
<td>Oranjetipje</td>
</tr>
<tr>
<td>Phylum</td>
<td>Arthropoda</td>
</tr>
<tr>
<td>ScientificName</td>
<td>Anthocaris cardamines (Linnaeus, 1758)</td>
</tr>
<tr>
<td>ScientificNameAuthorship</td>
<td>(Linnaeus, 1758)</td>
</tr>
<tr>
<td>SpecificEpithet</td>
<td>cardamines</td>
</tr>
</tbody>
</table>

Cees Hof
Netherlands Biodiversity Information Facility (NLBIF)

PIRE workshop 2015
Integrated Publishing Toolkit

See examples at:

http://ipt.nlbf.nl
http://www.cgn.ipt.wur.nl/
http://ipt.ibed.uva.nl
http://ipt.nioz.nl
http://ipt.ibed.uva.nl

Metadata Authoring
Primary Biodiversity Data
Species Checklists

The data publishing tool!

Cees Hof
Netherlands Biodiversity Information Facility (NLBIF)
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

Cees Hof
Netherlands Biodiversity Information Facility (NLBIF)
The data in the Toolkit of the dataprovider…
The Portal

http://www.gbif.org

Cees Hof
Netherlands Biodiversity Information Facility (NLBIF)
The data in GBIF

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

Summary

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

DESCRIPTION
The Limnldata 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; 25 water boards, the provinces and the Ministry of Transport, Public Works and Water Management. Over 25,000 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 COVERAGE
Date range: 1-jan-1980 - 31-dec-2010

LANGUAGE OF DATA

ADMINISTRATIVE CONTACT
Erie van der Wal

METADATA AUTHOR
Erie van der Wal

ORIGINATOR
Pieter Knoesbun

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.
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’.
More and more (smart) use of the GBIF services
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.
- 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.
- Remote sensing, infrastructure, urbanisation, crops, 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...

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...

Cees Hof
Netherlands Biodiversity Information Facility (NLBIF)

PIRE workshop 2015
http://www.gbif.org

http://www.nl bif.nl