The Global Biodiversity Information Facility (GBIF)

Metadata concepts and use in biodiversity

A presentation related to the e-infrastructure of the ESFRI project LifeWatch
In 2002, the world’s leaders agreed to achieve a significant reduction in the rate of biodiversity loss by 2010. Having reviewed all available evidence, including national reports submitted by Parties, this third edition of the Global Biodiversity Outlook concludes that the target has not been met. Moreover, the Outlook warns, the principal pressures leading to biodiversity loss are not just constant but are, in some cases, intensifying.

The consequences of this collective failure, if it is not quickly corrected, will be severe for us all. Biodiversity underpins the functioning of the ecosystems on which we depend for food and fresh water, health and recreation, and protection from natural disasters.

Cees Hof
Netherlands Biodiversity Information Facility (NLBIF)

MPI 8 September 2010
Biodiversity crisis.....

These graphs help to summarize the message from the available indicators on biodiversity: that the state of biodiversity is declining, the pressures upon it are increasing, and the benefits derived by humans from biodiversity are diminishing, but that the responses to address its loss are increasing. They reinforce the conclusion that the 2010 biodiversity target has not been met.

Most indicators of the state of biodiversity show negative trends, with no significant reduction in the rate of decline.

There is no evidence of a slowing in the increase of pressures upon biodiversity, based on the trend shown by indicators of humanity’s ecological footprint, nitrogen deposition, alien species introductions, overexploited fish stocks and the impact of climate change on biodiversity.

The limited indicators of the benefits derived by humans from biodiversity also show negative trends.
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.

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.

Priorities, with an emphasis on promoting participation and working through partners:

- include mobilising biodiversity data
- developing protocols and standards to ensure scientific integrity and interoperability
- building an informatics architecture to allow the interlinking of diverse data types from disparate sources
- promoting capacity building and catalysing development of analytical tools for improved decision-making
GBIF and LifeWatch: LifeWatch building on the infrastructure of GBIF, MoC in place.
Dividing the tasks...
GBIF focus:

Ecosystems:

Species:

Genetic:

Cees Hof
Netherlands Biodiversity Information Facility (NLBIF)

MPI 8 september 2010
GBIF data:

Focus on Primary Biodiversity Data….

Which species, where, when, by whom (occurrence records)

Additional:

- Names
- Multi media
- Metadata on occurrence records
GBIF infrastructure

Through appropriate standards and tools, the infrastructure is designed to serve:

- Metadata
- Primary biodiversity data
- Names data

Infrastructural components:

- Publishing
- Discovering
- Indexing
- Integrating
- Retrieving
- Analysing
GBIF facts and figures:

Distributed network of datasets
- 54 countries, 44 NGO’s
- > 200.000.000 records
- > 10.000 data sets

Access through GBIF portal:

http://data.gbif.org/
Example of a country summary page - gives the distribution and density of available records.
Functionality allows user to perform complex queries on the data.

In this example accessed all records on Amphibia in Tanzania from 1950 to 2007.
Challenges.....

Target: 1 bilj. Primary data    Target: 5 bilj. Through meta data

Cees Hof
Netherlands Biodiversity Information Facility (NLBIF)
GBIF data flow:

(Data digitization) → Data standardisation → Data publication → Data indexing → Data presentation

Red arrows indicate bottlenecks
GBIF core activity:

Data standardisation!

Darwin Core (DwC):

Darwin Core is a Biodiversity informatics data standard that consists of a vocabulary of terms to facilitate the discovery, retrieval, and integration of information about organisms, their spatiotemporal occurrence, and the supporting evidence housed in biological collections.

The Darwin Core is based on the standards developed by the Dublin Core Metadata Initiative and should be viewed as an extension of the Dublin Core for biodiversity information. The purpose of these terms is to facilitate data sharing by providing a well-defined standard core vocabulary in a flexible framework to minimize the barriers to adoption and to maximize reusability.

More than 80 fields, extensions in use, GBIF only needs a minimum of 6 core fields……
- Core taxon file (taxa.txt)
  …… A minimal requirement is that each row contains a unique taxonID.

- Extension files

- Dataset metadata
  An additional file describing the dataset as a whole, not pictured in the diagram above, may also be included in a Darwin Core archive, or optionally referenced in the Meta file by a URI. This file provides information about the entire published checklist such as the title of the checklist, authors, web and publication, information, etc.

  This information is stored either in the Ecological Markup Language (EML) or as simple Dublin Core (DC) as an xml file which is referenced via the metadata attribute of the archive element, e.g. `<archive metadata="eml.xml">`. 

Cees Hof
Netherlands Biodiversity Information Facility (NLBIF)

MPI 8 september 2010
Through appropriate standards and tools, the infrastructure is designed to serve:

- Metadata
- Primary biodiversity data
- Names data

Infrastructural components:

- Publishing
- Discovering
- Indexing
- Integrating
- Retrieving
- Analysing

GBIF uses tool for publication of primary data and metadata....
OAI-PMH

- Open Archives Initiative Protocol for Metadata Harvesting
- Providing a low-barrier mechanism for interoperability across distributed metadata repositories
- Data providers expose metadata; Service providers consume metadata through a client application known as a harvester that issues OAI-PMH service requests over HTTP:
  1. GetRecord
  2. Identify
  3. ListIdentifiers
  4. ListMetadataFormats
  5. ListRecords
  6. ListSets

1. return individual record  2. retrieve information about repository  3. retrieve headers of records
4. return metadata formats available  5. return records from repository  6. retrieve set structure (groupings) of repository

GBIF metadata access…..

GBIF: role as harvester and producer

http://www.openarchives.org/pmh/
GBIF (new) metadata infrastructure

Architecture: main components

A central metadata catalogue integrated in portal

A web based interface for searching/browsing

Many catalogues in network contributing metadata

A registry to manage network entities

Standards based protocol for communications

Tools for editing metadata

Goal:
A gateway to enable discovery and re-use of data on the GBIF and other biodiversity networks
GBIF (new) metadata infrastructure
Metadata Standards

Ecological Metadata Language (EML) v2.1.0
http://knb.ecoinformatics.org/software/eml/

Dublin Core (http://dublincore.org/documents/dcmi-terms/)

Directory Interchange Format (DIF)
http://gcmd.nasa.gov/User/difguide/difman.html

ISO 19115/19139 Geographic Metadata

Natural Collections Descriptions
http://www.tdwg.org/standards/312/

Federal Geographic Data Committee Biological Profile

Multimedia Resources Metadata Schema
http://www.tdwg.org/charters/article/view/448/36