GLOBAL BIODIVERSITY

INFORMATION FACILITY

Data Documentation, Quality, Preservation, and IPR—GBIF's Perspective
Jim Edwards
Executive Secretary, GBIF

Global Spatial Data and Information User Workshop
Lamont-Doherty Earth Observatory, Palisades, NY
21-23 September, 2004

GBIF.ORG and GBIF.NET

Global Biodiversity Information Facility

GBIF's goals are to …

- … Make the world's biodiversity data <u>freely and universally available</u> via the Internet

- … Share <u>primary scientific biodiversity data</u>
  - Especially <u>georeferenced</u> data

- … Promote the development of <u>biodiversity informatics</u> around the world

Global Biodiversity Information Facility

What do we mean by primary biodiversity data?

- <u>Label data</u> on ~ 1.5 - 3.0 billion specimens in natural history collections, herbaria, botanical gardens, etc.
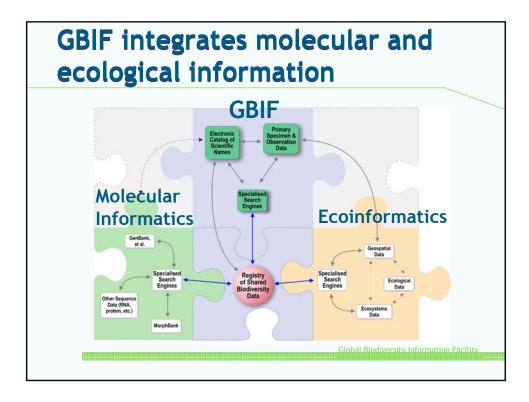

- Associated notes, recordings, publications, etc.
- <u>Observational data</u> (e.g. bird banding data)
- These data have been amassed over ~ 300 years; <u>most not digital</u>

GBIF uses a web-services approach

- Uses open-source software
- Data providers make their data known through our Registry of Shared Biodiversity data
  - <u>www.gbif.net</u> currently serving > 41 million specimen and observation records from > 250 collections
- Electronic catalogue of scientific names (ECAT) will be available as an authority file to any user
- Offer two simple mapping options developed by some of our Participants

GBIF integrates molecular and ecological information

Data Documentation

- Work closely with Taxonomic Databases Working Group (TDWG) to develop standards
  - For data descriptors
    - Support DiGIR and ABCD protocols
  - For taxonomic names
    - Are supporting development of a new Names Standard ("Napier Standard")
  - For geospatial data

Georeferenced Data

- Most new data use the recommended standards
  - Including GPS information
- Legacy data are a gigantic problem
  - Helping to develop automated tools
    - E.g. Biogeomancer
  - Support community approaches
    - E.g. MANIS, HerpNet
  - Data cleansing tools to identify outliers

Electronic Catalogue of Scientific Names (ECAT)

- ECAT is essential for linking data across all of biology
- It is estimated that about 1.8 million species have been given scientific names
- Many species have been described more than once
  - Average species has at least three synonyms
- ECAT working with other initiatives to complete the catalogue by 2010
- Taxonomists are known for being slow and inefficient
  - Looking for ways to make the taxonomic process more efficient

Need globally unique identifiers

- Currently is difficult to identify duplicated data
- Need to be able to distinguish collections, specimens, names, taxonomic concepts, literature references, etc.
- GBIF is therefore considering offering a GUID service to the biological community
  - At present find LSID best approach
  - Will be holding workshops over next year to gain community support for this idea

Intellectual Property Rights

- GBIF claims no IPR on data served through its portal
- Control of data remains with data providers
  - Each provider decides which data to serve
  - Sets its own policy for data re-use
- Most providers do not allow commercial use of data
  - But they differ in what this means to them
- Is a clear need for us to deal better with IPR issues
  - Have held one IPR workshop; plan others

How to contact GBIF

Web site: www.gbif.org
Data portal: www.gbif.net

GBIF Secretariat
Universitetsparken 15
2100 Copenhagen
Denmark

E-mail: gbif@gbif.org
Phone: +45 3532 1470
Fax: +45 3532 1480

New GBIF Secretariat headquarters, supported by
grant from Aage V. Jensens Fonde

Global Biodiversity Information Facility