

The future of biodiversity informatics: GBIF, the Encyclopedia of Life and beyond



Jim Edwards
Executive Director, EOL

GBIF Science Symposium
Amsterdam
18 October 2007

15 years of “biodiversity informatics”

- “Biodiversity informatics” first used in 1992
 - See historical description at www.bgbm.org/BioDivInf/TheTerm.htm
 - Coined by John Whiting, Canadian Biodiversity Informatics Consortium
 - Intended to aid business aspects of implementing the Convention on Biological Diversity
- How has biodiversity informatics developed since then?
- How is it likely to develop in the future?

Definitions of biodiversity informatics (BDI)--1

- Quick survey found more than ten different definitions
 - I’ll discuss six of them
 - 1. Whiting (1992)--merging of GIS, GPS, database management, environmental economics, museum catalogues
 - 2. OECD Working Group (1999)--application of informatics to recorded and yet-to-be-discovered information specifically about biodiversity, and the linking of this information with genomic, geospatial and other biological and non-biological datasets

Definitions of biodiversity informatics (BDI)--2

3. Berendsohn (2001)
 - The application of IT tools and technology to biodiversity information, particularly at the organismic level.
 - It thus deals with information capture, storage, provision, retrieval, and analysis, focused on individual organisms, populations, and taxa, and their interaction.
 - It covers the information generated by the fields of systematics (including molecular systematics), evolutionary biology, population biology, behavioural sciences, and synecological fields ranging from population biology to parasitism and phytosociology.

Definitions of biodiversity informatics (BDI)--3

4. Soberon and Peterson (2004)--application of information technologies to the management, algorithmic exploration, analysis and interpretation of primary data regarding life, particularly at the species level of organization
5. Heidorn (2007)--the study of data problems where information acquisition, analysis, sharing, and collaboration are required to answer broad questions about biodiversity
6. Commonwealth of Pennsylvania--the *sharing* of information in any format pertaining to the taxonomy, ecology, and genetics of organisms and biological communities

Creative ambiguity

- In his important 1989 book, *The Evolution of Phylogenetic Systematics*, David Hull argued that creative ambiguity about the definition of “species” was of great value and utility to systematics
- I would argue the same thing regarding biodiversity informatics
 - “we know it when we see it”

Some early BDI activities

- Environmental Resources Information Network (ERIN) (Australia)--late 1980s
- INBio (Costa Rica)--1989
- Conabio (Mexico)--1992
- Berlin Botanical Garden and Museum, Biodiversity Informatics Section--1992
- TDWG (Biodiversity Information Standards)--1985
- Integrated Taxonomic Information System (ITIS)--1996
- Species 2000--1996

Precipitating Event--GBIF

- OECD Megascience Forum Working Group on Biological Informatics
 - Began in 1996
 - Final report (1999) recommended establishment of GBIF
- GBIF--2001
 - Portal
 - Millions --> billions of data records
 - Standards
 - “Professionalize” TDWG
 - Nodes
 - National, regional, thematic
 - Ebbe Nielsen Prize
 - Science symposia
 - National, regional and organizational members

Period of explosive growth

- NCL Centre for Biodiversity Informatics (India)--2000
- Speciation event: *Biodiversity Informatics*--2004
- Ocean Biodiversity Informatics conferences--2004, 2007
- Species-bases sites: FishBase, AntWeb, AmphibiaWeb, North American Mammals, Swedish ArtDatabanken, Atlas of Living Australia, Netherlands species compendium ...
- Specimen-based networks: HerpNet, MANIS, ORNIS,
- Regional networks: IABIN, OBIS, ...
- Biogeomancer--2005
- IdentifyLife--2005
- JRS Biodiversity Foundation--2005
- European Distributed Institute of Taxonomy (EDIT)--2006
- BDI curricula
 - University of Illinois Master of Science in Biological Informatics--2006
- Encyclopedia of Life (EOL)--2007

How will we bring all these activities together

- Need to overcome “acronym fatigue”
- Find a way to interoperate these disparate organizations/movements/institutions
- One way is by adopting Web 2.0 technologies

The future--Web 2.0

- Wikipedia definition:
 - “a perceived second [generation](#) of web-based communities and [hosted services](#)--such as [social-networking sites](#), [wikis](#) and [folksonomies](#)--which aim to facilitate collaboration and sharing between users”
- Examples:
 - Wikipedia
 - Flickr
 - FaceBook

Web 2.0 tools

- Mashups
 - Wikipedia definition:
 - “In Jamaican creole, an exceptional performance or event”
 - “A web application that combines data and/or functionality from more than one source into a single integrated tool ... thereby creating a new and distinct web service that was not originally envisioned by [the contributing sources]”
 - Examples:
 - Rod Page’s iSpecies
 - Google Earth and its GBIF implementation

Web 2.0 tools

- Mashups
- Wikis--"computer software that allows users to easily edit, create, and link web pages. Wikis are often used to create collaborative websites"
 - Examples:
 - Wikipedia
 - GBIF
 - EDIT scratchpads

Web 2.0 tools

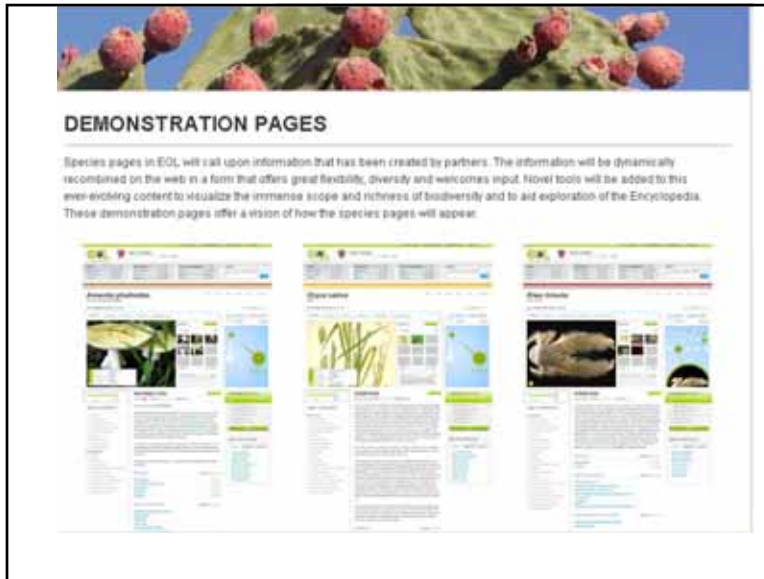
- Mashups
- Wikis
- Blogs
- RSS feeds
- Tagging
- Social networking

An example: The Encyclopedia of Life (EOL)

- An online encyclopedia composed of 1.8 million web sites
 - One for each known species
- EOL is developing two aspects of the original GBIF work programme
 - SpeciesBank--assemblage of all kinds of information about species
 - Digital library of biodiversity literature

Web 2.0 components of the Encyclopedia of Life (EOL)

- Each site consists of several components
 - Species page for the general public
 - Draft pages assembled via mashup technology
 - Drafts authenticated by experts ("curators") using controlled wikis
 - Information protected from being changed by anyone except the curators
 - But anyone can comment on the information and or suggest things to add
 - Curators will examine these suggestions and may move some of the information to the protected part



- ### Web 2.0 components of the Encyclopedia of Life (EOL)
- Each site consists of several components
 - Species page for the general public
 - Community-assembled spaces
 - E.g. taxonomists, molecular biologists, horticulturists, birdwatchers, pollinator biologists, etc., etc.
 - Each links in different databases and information
 - Can also be the focus of social networks
 - Spider freaks, leech aficionados, polar bear lovers, ginkgo groupies, microbe mavens, whatever ...
 - Each group/network controls the information on its space

- ### Example of a science-based community-assembled space on the EOL
- Scientists working on ageing wanted access to longevity information on the EOL
 - Proposed to organize their community to find this information and put it on the EOL species pages
 - Will set up their own portal into this information and manage the changing of the information
 - Received USD 2 million from private foundation to fund this activity

Example of an education-based community-assembled space on the EOL

- A school wishes to catalogue the biodiversity of a site near their schoolyard
- EOL and GBIF supply a bioblitz tool for them to use
 - Use GPS-enabled phones to take pictures of organisms found on the site
 - Assembly software combines these into a community inventory
- Students identify the organisms using EOL species pages
- Prepare inventory of the site
- Serve that information back to the EOL web pages (and potentially even to GBIF)

Web 2.0 components of the Encyclopedia of Life (EOL)

- Each site consists of several components
 - Species page for the general public
 - Community-assembled spaces
- Digitized biodiversity literature
 - Biodiversity Heritage Library--consortium of 10 of the largest natural history libraries
 - Scanning and marking up of 320,000,000 pages of literature

Biodiversity Heritage Library

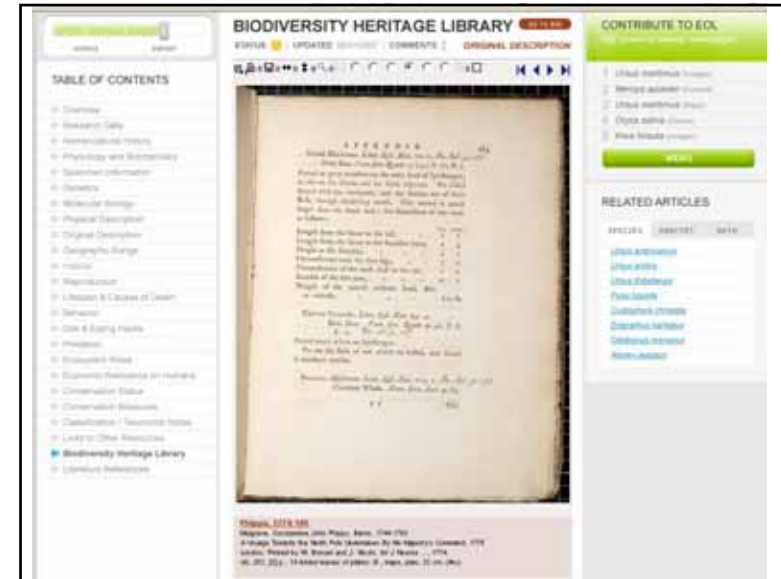


- American Museum of Natural History (New York)
- Field Museum (Chicago)
- Natural History Museum (London)
- Smithsonian Institution (Washington)
- Missouri Botanical Garden (St. Louis)
- New York Botanical Garden (New York)
- Royal Botanic Garden, Kew
- Botany Libraries, Harvard University
- Ernst Meyer Library of the Museum of Comparative Zoology, Harvard University
- Marine Biological Laboratory / Woods Hole Oceanographic Institution

Internet Archive Scribe: Boston



Internet Archive Scribe: London



Web 2.0 components of the Encyclopedia of Life (EOL)

- Each site consists of several components
 - Species page for the general public
 - Community-assembled spaces
- Digitized biodiversity literature
- Community tagging of information
 - No database can be prepared for all the questions users will want to ask
 - So let the users tag information in ways that will let them answer those questions
 - Similar to tagging on Flickr or Amazon

Web 2.0 components of the Encyclopedia of Life (EOL)

- Each site consists of several components
 - Species page for the general public
 - Community-assembled spaces
- Digitized biodiversity literature
- Community tagging of information
- RSS feeds
 - Users can ask for notifications whenever
 - A new species page is added to a taxon
 - New information is added to a species page
 - Newly published literature about a taxon is available
 - Newly scanned literature about a taxon is available on the BHL

Web 2.0 components of the Encyclopedia of Life (EOL)

- Each site consists of several components
 - Species page for the general public
 - Community-assembled spaces
- Digitized biodiversity literature
- Community tagging of information
- RSS feeds
- Blogs
 - EOL looking for bloggers
 - Scientific
 - Social
 - For/by young persons

Speed bumps

- Sustainable funding for databases
- Lack of globally unique identifiers (GUIDs)
- Continued funding for underlying science
- Money for content development
- Plethora of initiatives
 - Haven't yet coalesced into a synergistic whole
- Need to continue to expand people involved
 - Same people keep getting overworked
- Training and infrastructure, especially in developing countries

What next for BDI?

- Prognosis is good
 - Google search on “biodiversity informatics” returns about 114,000 hits
 - More being added every week
- Integration of BDI activities and initiatives
 - Len Hirsch--show how BDI initiatives link together to form a strong edifice
 - International biodiversity informatics meeting--2009
- Greater involvement of general public in using and providing information
 - Increased use of Web 2.0 thinking
- Explosive use of BDI information
 - Recognition of BDI as a “public good”

Tipping point for BDI

- Are really beginning to see the truly seamless biodiversity information environment that was envisioned by the OECD Working Group:
 - “Correlating information from disparate sectors of knowledge”
 - “Unlock[ing] the knowledge and economic power lying dormant in the masses of biodiversity data that we have on hand” (and that are currently being gathered)
 - Putting into place an “ecosystem” of interoperating biodiversity databases and resources that serves an expanding base of users

In a nutshell,

We're entering in to the golden age of
biodiversity informatics

Let's make the best of it!

Thank yous

- Three people have played an important role in helping me to learn about the power and potential of BDI
 - Meredith Lane
 - Donald Hobern
 - Rod Page