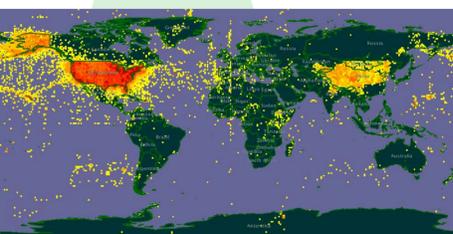


GBIF

www.gbif.org

...free and open access
to biodiversity data

The GBIF Data Portal: a gateway to biodiversity information



Uncorrected occurrence data for records located in the USA. Note 'shadow' of the USA over Asia, due to incorrect longitude entries.



Occurrences for the USA as seen from mid-2011, following processing improvements



Occurrences for the USA as seen from late 2011, following further improvements. Note coastal zones are now included.

Background

The GBIF Data Portal has provided discovery and access services to the content shared through GBIF since its launch in 2007. A key resource accessible through the portal is primary species occurrence data such as specimens in museums, observations in the field and culture strains.

Issues

The primary critiques received of the portal relate to the quality of content and the time lag between data being added or changed on the network, and becoming visible on the portal. Other areas identified for improvement were the backbone taxonomy used to organize the content, which had become disorderly as records grew rapidly in number; and the geographic and temporal interpretation of records (see maps in left column).

Solutions – improvements to the portal in 2011

During 2011, GBIF has developed new processing workflows to include the following:

- reduced processing time for data from 3-4 days to around 36 hrs, enabling more frequent rollovers (publication cycles) and therefore shorter delay between entering data and appearance on portal;
- a complete regeneration of the taxonomic backbone, including up-to-date taxonomic catalogues;
- enhancements of the registry (see poster on GBIF Registry) to support better modelling of the GBIF network structure, players and interactions. This paves the way for better attribution of data owners and service providers; and
- extended data interpretation routines and plausibility checks, e.g. in the areas of geolocation, date interpretation, altitude and depth, etc. (for illustration of some results, see maps in left column).

The solutions were developed to run on the Hadoop platform; an efficient parallel processing environment enabling the GBIF Portal to grow with future data volumes.

Looking ahead – plans for 2012

By the end of 2012, GBIF aims to add the following key functionalities to the data portal:

- Indexing of metadata networks to provide discovery of undigitized content and content shared in non-standard forms (see poster on metadata)
- Scientific name indexing of journal publication feeds (uses name finding services such as TaxonFinder – see poster)
- Improved means to determine the fitness for use of content accessible in the GBIF network
- Means for end users to annotate content and help improve data quality
- More accurate citation of data publishers
- Indexing of a broader range of content types, such as sound records, images and videos (see poster on Audubon Core)

About GBIF

The Global Biodiversity Information Facility (GBIF) was established by governments in 2001 to encourage free and open access to biodiversity data, via the Internet. Through a global network of national and thematic nodes, and a Secretariat based in Copenhagen, Denmark, GBIF promotes and facilitates the mobilization, access, discovery and use of information about the occurrence of organisms over time and across the planet.

Vision - A world in which biodiversity information is freely and universally available for science, society, and a sustainable future.

Mission - To be the foremost global resource for biodiversity information, and engender smart solutions for environmental and human well-being.

www.gbif.org

