GBIF Integrated Publishing Toolkit (IPT)

Publishing biodiversity data to the world

The IPT allows for efficient and easy sharing and hosting of organism occurrence data, taxonomic and nomenclatural information, and general dataset metadata. By focusing on specific biodiversity data types, the software can provide a richer environment than current generic wrapper solutions like DIGIR² provide.

Key features
- data cache
- managed record metadata
- rich web application
- OGC Web Feature Service²
- TAPIR lite³
- Ecological Metadata Language⁴ metadata
- Java (Struts2/Hibernate/Spring)⁵
- user accounts / roles

The Cache

An embedded or external database is used as a cache that serves the IPT user and web service interfaces. Existing relational databases may be connected or files in delimited format (e.g. tab delimited) may be imported during which GUIDs may be assigned and changes or deletions of records are detected, a feature needed by incremental harvesting protocols, e.g. OAI-PMH⁶.

Each served data type has a core entity which extension tables can link to in a one-to-many relation. Extensions can dynamically be added to the cache as required by the administrator and suitable extensions can be discovered in a shared, central repository.

Interfaces to data

A rich web application publishes data to the web, providing statistics & maps using the Google Chart API, and fulltext search with Lucene⁷. Each served data type has a core entity which extension tables can link to in a one-to-many relation. Extensions can dynamically be added to the cache as required by the administrator and suitable extensions can be discovered in a shared, central repository.

A compressed archive of tab delimited files containing all cached data is served for each dataset while incremental harvesting is supported via OAI-PMH. TAPIR lite is implemented for occurrence data based on Darwin Core⁸ and ABCD⁹ and simple REST¹⁰ services serve XML and JSON¹¹ records, including all extension data. Taxonomic checklists are preconfigured utilising Geoserver¹² that is also used to provide maps using the OGC WMS¹³ protocol.

General dataset metadata can be authored in the GBIF profile of EML and published to the GBIF metadata catalogue, and is also discoverable through a RSS¹⁴ feed covering the latest modified resources.

About GBIF

GBIF makes digital biodiversity data openly and freely available on the Internet for everyone, and endorses both open source software and open data access.

www.gbif.org

GBIF provides scientific biodiversity data for decision-making, research endeavours and public use.
data.gbif.org

GBIF is a network of data publishers who retain ownership and control of the data they share. Linked datasets provide a more robust representation of biodiversity than any single dataset.

GBIF provides access to primary biodiversity data held in institutions in developing and developed countries. Data shared through GBIF are repatriated data.

GBIF is a dynamic, growing partnership of countries, organisations, institutions and individuals working together to mobilise scientific biodiversity data.

GBIF invites you to download species occurrence data freely and openly from http://data.gbif.org

GBIF invites you to join the GBIF network and share your biodiversity data, as well as participate in developing new tools and services.

Project home:
http://code.google.com/p/gbif-providertoolkit/

1 http://digg.sourceforge.net/
2 http://www.opengeospatial.org/standards/wfs
3 http://www.tdwg.org/dav/subgroups/tapir/1.0/docs/
4 http://knb.ecoinformatics.org/software/eml/
5 http://struts.apache.org/2.x/
6 http://www.hibernate.org/
7 http://www.csrf.org/
8 http://www.rssboard.org/rss-specification
9 http://www.opengeospatial.org/standards/wms
10 http://geoserver.org
11 http://www.springframework.org/about
12 http://www.hibernate.org/
13 http://www.opengeospatial.org/standards/wfs
14 http://www.jboss.org/announcements