



Project no. 018340

Project acronym: EDIT

Project title: Toward the European Distributed Institute of Taxonomy

Instrument: Network of Excellence

Thematic Priority: Sub-Priority 1.1.6.3: "Global Change and Ecosystems"

C5.89 List of data items and searchable data items in existing portals

Due date of component: Month 26

Actual submission date: Month 26

Start date of project: 01/03/2006

Duration: 5 years

Organisation name of lead contractor for this component: 9 FUB-BGBM

Revision [draft 1]

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

C5.89 List of data items and searchable data items in existing portals

Compile the specimen access functionality of existing specimen access portals for zoological, mycological and botanical specimens. Attempt to develop a priority list of access points.

Examine existing specimen access portals for zoological, mycological and botanical specimens.

In April 2008, we examined the structure, search and filter functionality of 64 data portals and online specimen databases. These represented either institutions offering Internet access to specimen data in their own collections, or scientific associations with a specific focus, i.e. specimens of a particular higher taxon or ecological group. Databases and portals accessing such specialised collections were of particular interest, because they allowed us to study features relevant to specific subjects, and to study the specific needs of biologists working in different taxonomic groups or areas. To find specific and non-specific databases of zoological, mycological and botanical collections, we mainly used targeted Google queries.

Results

50 of the 64 Internet specimen databases we examined are restricted to one collection, one institution or one topic (e.g. types); 14 provide specimen data from several specimen collections. Only two of them provide live data, with only a few basic search functions and slow query response times.

The Internet portals under examination offer a wide range of access points and possible queries:

- genus
- species
- infraspecific taxa
- taxon name (string)
- higher rank
- results limited to one higher taxon
- host
- author of taxon
- year published
- current name
- common name
- collector / contributor (images)
- collector's number
- date / collecting period
- collecting event
- project name
- determined by (when)
- collection
- institution
- barcode
- accession number
- catalogue number
- country
- locality / region
- coordinate range
- elevation
- habitat
- type

- type status
- preparation type
- sex/caste
- life stage
- GenBank number (NCBI)
- age
- element
- formation
- image type
- image metadata
- publication
- cultivation (country, land district, ecological district, time)
- donor institution
- duplicate to
- Exsiccatae
- red list status

Search functions:

- include synonyms
- filter function for results
- records with image only
- records with sound only
- records with flower / fruit only
- sort function for results
- administrative database info
- number of records displayed
- keyword query
- further search functionality

An overview of access points is given in Fig. 1:

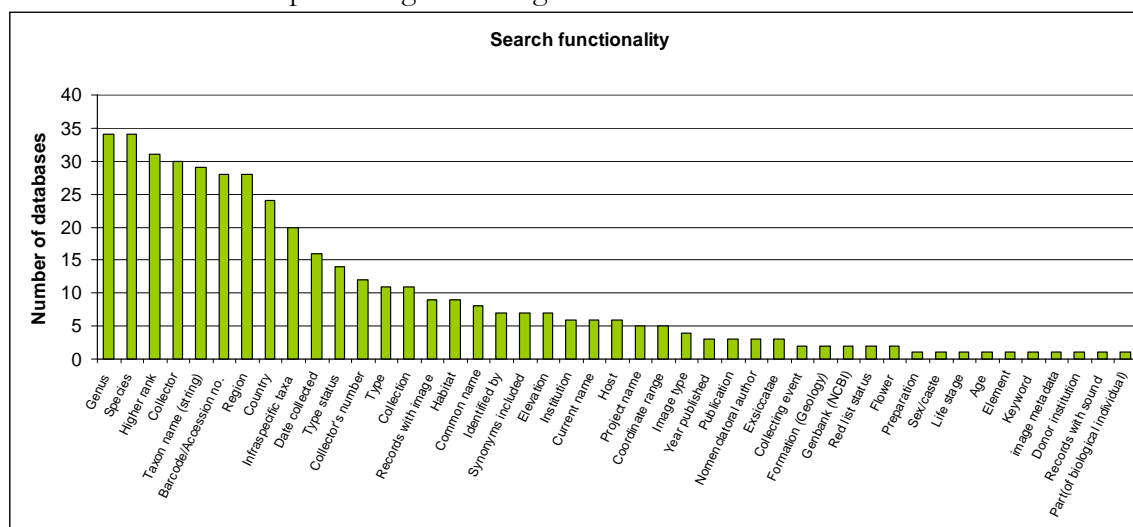


Fig.1. Access points of data portals and specimen databases by number.

In biological online specimen databases, the main data access point is species or genus name, and occasionally infraspecific taxon name. In addition, specimens are usually searchable via collector name, barcode or accession no.; the region or the country in which the specimen was collected; the date it was collected; type and type status; collector's number, collection or institution; habitat; common names; identifier elevation; current name; and host of the specimen. A few databases offer the inclusion of synonyms in the search or the filtering of results to those with digitized images only.

The BioCASE interface contains most of these access points, especially the most important and commonly used. Taxonomists would prefer additional access via real collection and institution, barcode, host, habitat and substrate, as well as via more specialized areas such as sex, stage and preparation/fixing.

Although BioCASE covers the needs of users in most biological disciplines, the requirements of highly specialized fields like palaeontology make it impossible to design a completely generic portal. For meaningful access to paleontological databases, a separate portal must be implemented.

Example databases checked:

- ANHSIR: ANH Specimen Information Register [<http://www.anbg.gov.au/cgi-bin/anhsir>]
- Botany Types Databases [<http://www.bishopmuseum.org/research/natsci/botany/dbandkeys/botanydb.html>]
- Diatom Collection Hanna Database [http://research.calacademy.org/research/diatoms/hanna_db/index.asp]
- Essig Museum of Entomology [<http://essigdb.berkeley.edu/advanced.html>]
- GBIF [<http://data.gbif.org>]
- GBIF-Botanik Prototype [<http://www.gbif.de/botanik/datenabfrage/default.html>]
- Royal Botanic Garden Edinburgh [<http://rbg-web2.rbge.org.uk/multisite/multisite3.php>]
- Southwest Environmental Information Network [<http://seinet.asu.edu/seinet/collections/index.jsp>]
- SYNTAX [http://www.biologie.uni-ulm.de/syntax/portal/index_d.html]
- The Collection of Lichenicolous Fungi at the Botanische Staatssammlung München [http://www.botanischestaatssammlung.de/DatabaseClients/BSMlichfungicoll/DiversityCollection_BSMlichfungicoll_Find.cfm]
- Tropicos [<http://www.tropicos.org/NameSearch.aspx>]