From the World Directory of Collections of Cultures of Microorganisms towards Mash-up of Biodiversity and Sequence Databases

Director, WFCC-MIRCEN World Data Centre for Microorganisms

V ice-chair, Governing Board of Global Biodiversity Information Facility (GBIF)

Center for Information Biology and DNA Data Bank of Japan (DDBJ)

National Institute of Genetics

Hideaki Sugawara
Evolution of Culture Collections to BRC

Global BRC Network (GBRCN) for harmonization, efficiency and sustainability

OECD BRC report
- Underpinning the future of life sciences and biotechnology

Quality of services - accreditation/certification

MTA/IPR

Biosecurity

Legislation

Biodiversity

Genomic

Informatics

Culture collections (low sustainability)

The World Directory to federate culture collections
Prototypes of information network for GBRCN.

- Global networks, e.g.
  - WFCC-MIRCEN* World Data Centre for Microorganisms (WDCM)
  - Global Biodiversity Information Facility (GBIF)
- Regional networks, e.g. EBRCN
- National networks, e.g. CCCM, JSCC, TNCC, UKNCC

* WFCC: World Federation for Culture Collections
MIRCEN: Microbial Resources Centers Network
World Federation for Culture Collections

- Annual Conference of the Association for General and Applied Microbiology (2006/3/19-22) [PDF] [Word] uploaded!
- EBRCEN transport paper WP5 Revision 13Jun05.doc uploaded!
- The Biology of Vibrio: Biodiversity, Genomics, Disease/Epidemiology, Ecology, and Applications, Ghent, Belgium (2005/11/07-08) uploaded!
- VII International Conference on Environmental Pollution (2005/09/20-25)
- Congress of PERMI and ACLA-B-3 (2005/09/25-27)
- What's New

About WFCC
- Please join us (WFCC)
- WFCC Executive Board
- WFCC Task Forces and Activities
- WFCC Statutes
- WFCC Bylaws
- Brochures
- Messages/Proposals from WFCC Executive Board
- WFCC Newsletter
- Publications on culture collections and microbial resource centers
- WFCC Workshops and Conferences
- Documents on bioterrorism and biowarfare
- Calendar of Events relevant to microbes and more

We are here to help collections in difficulty (Please visit What's New)
- WFCC membership application
- Contact us!

- Home Pages of Culture Collections in the world
- Statistics on Culture Collections in the world
- Federations, Societies and Networks of Microbial Resource Centers
- WFCC-MIRCECN World Data Centre for Microorganisms

http://www.wfcc.info/

501 culture collections in 66 countries
175 culture collections publish catalogues
154 culture collections have Web pages
WFCC-MIRCEN World Data Centre for Microorganisms (WDCM) provides a comprehensive directory of culture collections, databases on microbes and cell lines, and the gateway to biodiversity, molecular biology and genome projects.

SEARCH ENGINE
- Simple Search
- Search by Fields
- AHMIT
- Lineage search NEW

REGISTRATION/UPDATE
- CCINFO
- STRAIN

Please search menu(s) in this site by entering keyword in the window below

Global search by [Google]

Contact us!

Home Page in Japanese

http://www.wdcm.org/

Registration of culture collections and their holdings

WDCM assigns culture collections, cf. accession number of sequence data assigned by INSDC

Search engines
WDCN wrapped several databases for one-stop query for a search engine named AHMII.

A parallel access to multiple data sources in the Internet

AHMII call CGIs of the member Web sites.
AHMII requires neither change nor expansion of the member Web sites.
AHMII counts on query functions of the member Web sites.
AHMII does not display results in a standardized format.
WDCM has to maintain the program and the list of the member Web sites.

More sophisticated system than AHMII is now feasible.
Web services provide more sophisticated way to aggregate biological data resources

Web services provide human interface (Web browsing) and programmatic interface as well.
Aggregation of taxonomy, phenotypes, sequence data, environmental condition, location, references and even socio-economic data will be the base of the problem solving environment for R&D and decision makers.
Examples of mash-up: INSDC and GBIF

INSDC taxonomy DB
TxSearch Web service
INSDC sequence DB
SRS Web service

Specimen DB
GBIF Providers
Specimen DB
GBIF Providers
Specimen DB
GBIF Providers

Dynamic linkage of the sequence database and GBIF data providers

Option of the output
- without accession number of sequence data of INSD (default)
- with accession number of sequence data of INSD

The system will not display the list of accession numbers, if it takes more than 2 minutes to return the result.

It happens if the number of hits is comparatively large.

- Search
- GBIF Data Use Agreement

Scheme of the dynamic linkage

Web services by SOAP servers

TxSearch
Web service

SOAP

DiGIR

MySQL
Local Index

DiGIR (protocol)
(YAML)
More fine linkage of specimen including microorganisms cultures and sequence data by “structured vouchers”

- <institution-code> - abbreviation of the archiving institution
- <collection-code> - collection within the institution (possibly null)
- <specimen-id> - specimen identifier within the collection

/specimen_voucher="<institution-code>|<collection-code>|<specimen-id>"

- museums
- herbaria
- culture collections
- stock centers
- germplasm repositories (seed banks)
- frozen tissue banks
- zoos/aquaria/botanical gardens
- DNA banks
- personal collections
- e-voucher archives
Example of mash-up: GBIF and mapping tool
Example of mash-up: GBIF, data analysis tool and mapping tool

**Biodiversity Analysis Tool - Define Analysis**

- **Data Source**: Plants - Flora of Japan Specimen Database
- **Grid Size**: 1 Degree (e.g. 1 Degree)
- **Class**: Cycadopsida (e.g. Magnoliopsida)
- **Order**: Cycadales (e.g. Asterales)
- **Family**: Cycadaceae (e.g. Asterales)
- **Analysis Type**: Richness, Endemism, Taxonomic Diversity
- **Output**: MapServer, Google Earth

Run the Analysis  Reset

BAT has been implemented in NIG based on collaborative work with Australia.
With MapServer

Biodiversity Analysis Tool - Richness of Cycadopsida

- Data Source: Plants - Flora of Japan Specimen Database
- Grid Size: 1 Degree
Display richness of species in Google earth

Specimen data were from a plant data file in Japan GBIF node.
Display richness of species in Google earth

Specimen data were from a plant data file in Japan GBIF node.

Grids and color codes are displayed based on outputs from BAT system.
Summary

- The next generation WDCM will depend on a distributed system based on standard data schema and data transfer protocol.
- The question is who designs and maintains the standards.
- The next generation WDCM will provide a stable database that supports LSID (Life Science IDentifier) for the linkage of cultures and their data on genotypes and phenotypes.
- The question is who funds these activities for a long term.