

NETWORKING OF BIOLOGICAL RESOURCE CENTERS: WDCM EXPERIENCES

Satoru Miyazaki ^{*1} and Hideaki Sugawara ²

^{*}1 Center for Information Biology and DDBJ, National Institute of Genetics 1111 Yata Mishima Shizuoka, JAPAN

E-mail: smiyazak@genes.nig.ac.jp

2 Center for Information Biology and DDBJ, National Institute of Genetics 1111 Yata Mishima Shizuoka, JAPAN

E-mail: hsugawar@genes.nig.ac.jp

ABSTRACT

The WFCC-MIRCEN World Data Centre for Microorganisms (WDCM) was set up more than 30 years ago as a data center of the World Federation for Culture Collections (WFCC). It published the World Directory of Collections of Cultures of Microorganisms when it was established and now provides a portal site for microbial resource centers and their customers by fully utilizing Internet technology. This paper introduces international initiatives on biological resources centers together with the activities of WDCM.

Keywords: Database, Network, XML, Biological resource center, Microbiology, Biodiversity

1 Introduction

The WDCM is now the data center for the WFCC as well as MIRCEN (Microbial Resources Centers network).

On the WFCC Homepage ([WFCC](#), n.d.), you will find the following introduction to the WFCC under "About WFCC":

The WFCC is a Multidisciplinary Commission of the International Union of Biological Sciences (IUBS) and a Federation within the International Union of Microbiological Societies (IUMS). The WFCC is concerned with the collection, authentication, maintenance and distribution of cultures of microorganisms and cultured cells. It aims at promoting and supporting the establishment of culture collections and related services, to provide liaison and set up an information network between the collections and their users, to organize workshops and conferences, publications and newsletters and work to ensure the long term perpetuation of important collections.

MIRCEN is sponsored by UNESCO and is described at the Web page of UNESCO (UNESCO, n.d.) as:

MIRCEN is an acronym for Microbial Resources Centres that are actually existing academic and/or research institutes in developed and developing countries. These centres, in co-operation with the concerned National Commissions of Member States and governmental authorities, participate in a global collaborative network effort in the harnessing of the beneficial applications of the microbial world for human progress through the vehicle of international scientific co-operation.

Thus the WDCM is a vehicle for networking microbial resource centers of various types of microbes in developed and developing countries. It also serves as an information resource for the customers of the microbial resource centers. It should be noted that "culture collections" and "microbial resource centers" are different concepts according to the definition of the Organization for Economic Co-operation and Development's (OECD) Working Party for Biotechnology,. This issue will be touched upon in the Section 3 of this paper.

2 A short history of WDCM

The WFCC needed comprehensive information on culture collections world-wide to accomplish its mission. Among experts in the WFCC, Professor V. B. D. Skerman (University of Queensland, Australia) and his colleagues in the 1960's pioneered the development of an international database on culture resources worldwide. The result was the WFCC World Data Centre for Microorganisms (WDCM). He published the World Directory of Collections of Cultures of Microorganisms (the World Directory) in 1972 (Skerman, 1972), although the database itself was for in-house use only. The records in the CCINFO database contain data on the organization, management, services and scientific interests of the collections. Each record is linked to other types of records contained in the STRAIN database, which is a list of holdings of culture collections: algae, cyanobacteria, bacteria, fungi, yeasts, lichens, protozoa, tissue cultures and viruses.

The WFCC relocated the data center to the Institute of Physical and Chemical Research (RIKEN) Japan after an open competition to host the data center, when Professor Skerman retired. The WDCM in RIKEN was initially headed by Professor Kazuo Komagata and later by Dr. Hideaki Sugawara. The WDCM at RIKEN made the World Directory database open to the public via a Packet Switching System and then via the Internet using a Gopher system and then a Web site. In April, 1997, the WDCM moved from RIKEN to the National Institute of Genetics (NIG) after agreement among WFCC, RIKEN, NIG and Japan Society for Culture Collections (JSCC), when Dr. Hideaki Sugawara moved to the Center for Information Biology and DNA Data Bank of Japan, NIG. WDCM at NIG dropped the Gopher service and concentrated on providing Web-based services.

In 1998, the WDCM carried out a world-wide survey on culture collections sponsored by the Science and Technology Agency (STA) of the Japanese government. The CCINFO database was updated based on the results of this survey. In 2002, sponsored by UNESCO the WDCM again did the postal survey. This time, however, culture collections are able to respond in electronically either by E-mails or via a Web browser.

In addition to the mandatory task of maintaining the World Directory, WDCM researches and develops other databases and information systems. It carried out experiments on distant learning and virtual laboratories using the broad-band network between US and Japan (Atago, Kikuchi, Tateyama, Fujita, Harada, Yokokawa et al, 1997). It also developed the following tools (Sugawara, Miyazaki, Shimura & Ichianagi, 1996) :

- Agent to Help Microbial Information Integration (AHMII): to search multiple databases distributed on the Internet
- e-Workbench (named Information-base): to develop personal databases and carry out the polyphasic analysis of microbes based on phenotypic data and sequence data utilizing servers distributed over LANs and the Internet.

3 Networking of biological resource centers

Culture collections are required to meet the needs of mega-science projects such as biodiversity conservation and genome projects in the 21st century. Some international coordination is a prerequisite for fostering and networking these culture collections, because none can support such project requirements alone. At WDCM we organized a WDCM symposium on February 16th, 1999 in Tokyo to discuss cooperation between curators of culture collections worldwide in collaborative activities to meet demands from relevant scientific communities.

The symposium was followed by a closed workshop of the OECD Working Party for Biotechnology. The workshop invited experts from academia, industry and funding agencies to analyze and discuss the future of culture collections, in other words, the rationale for supporting culture collections by public funding. There was a consensus among the participants to upgrade conventional culture collections to the level of biological resource centers (BRCs) that could meet the challenges from genomics and biodiversity and also fully utilize informatics. Establishing a Task Force for BRCs closed the workshop. The Task Force was chaired by one of the authors (Hideaki Sugawara) and compiled a report entitled "Biological Resource Centres: Underpinning the Future of Life Sciences and Biotechnology" (OECD, 2001). The

report available on-line from the WDCM Web page (WDCM, 1997) calls for five actions by OECD countries and beyond:

- The establishment of national BRCs
- The development of an international accreditation system for BRCs
- The creation of international partnerships among BRCs
- Enhanced international co-ordination and the harmonization of standards, rules and regulations for BRCs
- The establishment of a global BRC network (GBN)

Information networks are definitely the infrastructure of GBN. The WDCM is a potential hub of the GBN, especially for microbial resource centers. The center is active in evaluating up-to-date information technologies to improve information networks just as it applied Gopher and the Web in 1990s. However, information networks per se are not enough for the networking of biological resource centers.

International partnerships for harmonizing standards etc. (see point 4 above), burden sharing, and exchange of experts and the dissemination of biological materials are all required for GBN. Information resources including the WDCM are part of the infrastructure of the international network.

4 Technologies and services in WDCM

WDCM has introduced XML technologies for data management, e.g. links in the WDCM Web pages are described as XML documents. This makes it possible to directly access data source(s) from hit(s) to the WDCM Web pages. It is also easy to sort by data items and select data items from the display of the search results. This is the power of XML tags.

XML is also used in WDCM for transactions of the relational database as shown in Figure 1. In the XML application server based on Cocoon, eXtensible Server Pages (XSP) creates an XML file in the computer memory when a query is made to the relational database (RDB). eXtensible Stylesheet Language Transformation (XSLT) transforms the XML file into a HTML file that any Web browser can display. XSL-FO develops a PDF file for printing or graphics for displaying statistics, for example. In this way, WDCM is able to disseminate data in various forms without writing conversion programs many times.

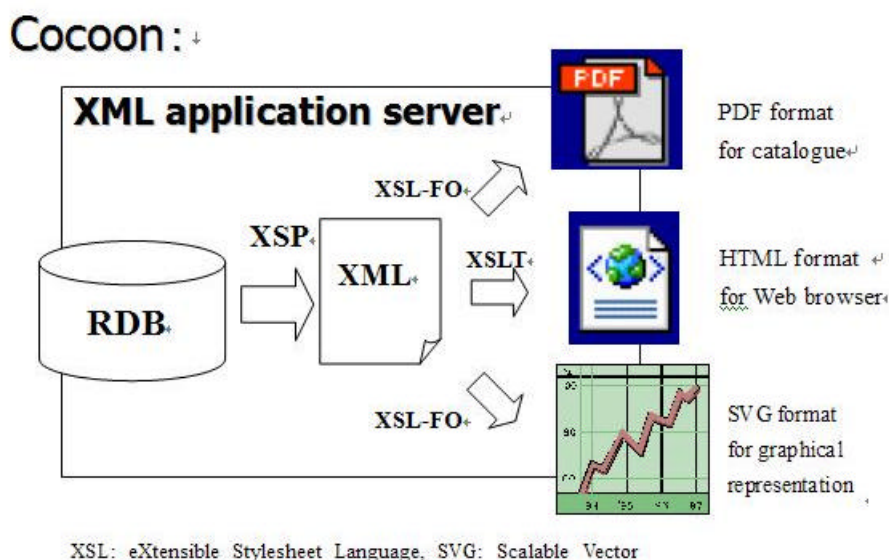


Figure 1 The XML application server based on Cocoon

To network the microbial resource centers, WDCM developed an on-line registration and updating system for the CCINFO and STRAIN databases in 2002. A center new to the CCINFO database is supposed to click the "Add" button to start entering their data in a sequence of Web pages. The center will be given a unique set of user-id and passwords to use when it updates the data in the database or when it wants to delete the registration. The categories of data items are: name, correspondent and addresses of the center, sponsors, personnel, main subjects, preservation of cultures, availability of

cultures, catalogue and services. Multiple-choice options are available for most of the data items to avoid typing errors.

It was obvious that information in the database would be more frequently updated after introducing the on-line system. However, ironically withdrawal from the database has also increased. Using the postal system, centers were more reluctant to reply to the survey than with the electronic system. As of May 2002 the CCINFO database includes 466 culture collections (CC) and microbial resource centers (MRC) in 61 countries, even though WDCM issued 824 registration numbers, i.e. 358 CC/MRC have disappeared from the directory. As far as WDCM is able to determine, these disappearances were mainly caused by either the retirement of the curator or funding shortages.

In the case of the STRAIN database, centers are able to upload the most up-to-date list from disks in their local machine or send files by E-mail.

Information collected by WDCM is disseminated from the Web site. The databases available with an integrated search system <http://www.wdcm.nig.ac.jp/cgi-bin/search.cgi> (WDCM, 1997) are summarized in Table 1. Users can check one or more database(s) to search simultaneously, e.g.: if you check "Bacteria Nomenclature" and "STRAIN bacteria", you will be able to check the authenticity of a scientific name and availability of strains all at once; you will be also guided to the CCINFO database as a result of searching for "STRAIN bacteria" to locate center(s) that maintain the strain.

The WDCM intends to implement a SOAP server and as well as Web Services Description Language (WSDL) in the near future as the authors did for DNA Data Bank of Japan (DDBJ, 2002)

5 Discussions and Conclusion

WDCM accomplished the mandatory task given to it by WFCC by publishing a hard copy of the World Directory, in 1970s and 1980s. The database was mainly for in-house use. In the early 1990s, the databases went on-line through a packet-switching system (PSS). However, this type of on-line access was rather expensive and complicated. Once the WDCM launched the Web site, the number of people accessing it increased to on average 55,000 hits/month in the year 2002. Accesses have been broken down into: abroad 14%, domestic 15%, commercial 39%, unknown 32%; Africa 1%, America 18%, Asia 13%, Europe 62% and Oceania 6%.

It is very important for a data center to evaluate emerging technologies to determine if they are powerful enough to deal with continuously expanding databases and evolving queries and also those that will become *de fact* standard technologies (e.g. the Internet and the Web). In a couple of years, XML will be standard in bioinformatics within the domains of genomics, proteomics and so on. It is true that XML will make our life easier than before. However, there is a risk that chaos caused by disparate technologies will still remain. In the meantime, heterogeneous Document Type Definition (DTD) and/or XML schema have already been defined and used for common biological objects and categories. The same heterogeneity will also be created in WSDL and UDDI too. The authors expect that "unity will be reached eventually" (Stein, 2002) after *natural selection*.

A data center supports a biological resource center and *vice versa*. An information system on its own is not enough for the data center to develop this bi-lateral partnership. It will be useful for the data center to organize meetings where participants are able to discuss issues related to the banking of biological materials and where they can also exchange relevant information and experiences, i.e. the data center should be a hub of the community in addition to being the hub of information flow.

The data center is expected to provide reliable data and therefore data cleansing is an issue. To some extent inconsistency among data entries may be solved by the data center. However, high-quality biological databases will be realized only if the data center is accepted and helped by the scientific community. The unification of scientific names from hundreds of centers is much more difficult than it looks. It can be done automatically to a certain level but can not be completed without experts, i.e. taxonomists. It is a role of the data center to develop an electronic dictionary and the tools for data

validation, to propose standardization of data structures and data representation and to guarantee the interoperability of information systems. On top of the efforts made by the center, these efforts have to be accepted by the community for the success. WDCM at Queensland prepared a form for the description of cultures including source, growth condition, and physiological, biochemical and genetic characteristics, Nevertheless, in practice the WDCM at NIG now only collects the names of cultures. Among the 498 centers registered in the database CCINFO, 78 centers publish catalogues and provide a Web site, that is, 78 centers explicitly intend to disseminate their information and have access to the Internet. So far the number of centers that have updated the CCINFO database on-line is not 78 but 58. In the case of nucleotide sequences, authors are obliged to register the data with either DNA Data Bank of Japan (DDBJ), EMBL database of European Bioinformatics Institute (EBI) or GenBank database of National Center for Biotechnology Information (NCBI) when author(s) submit a paper to a journal. It is now common for authors to submit sequence data, even if they do not write a paper. Generally though, scientists are not obliged to submit their data to most of the other biological databases.

In future, the WDCM aims to look into the role and value of the data center in an age when more and more centers will be online; strive to improve the understanding of microbial resources centers about the value of the data center; consult with such international initiatives of OECD GBN, Global Biodiversity Information Facility (GBIF, n.d.), Global Environmental Facility (GEF, n.d.) and many other international, regional and national efforts.

6 REFERENCES

Atago, T. Kikuchi, S., Tateyama, J., Fujita, M., Harada, H., Yokokawa, T., Sugawara, H. & Miyazaki, S. (1997). US-Japan cooperative experiments on broad-band network (in Japanese). *JOHO-KANRI* 40(5), 404-413

DDBJ (2002) Homepage of X-DDBJ project at DNA Data Bank of Japan. Available from: <http://xml.ddbj.nig.ac.jp/>

GBIF (n.d.) Homepage of Global Biodiversity Information Facility. Available from: <http://www.gbif.org/>

GEF (n.d.) Homepage of Global Environmental Facility. Available from: <http://www.undp.org/gef/>

OECD (2001) Biological Resource Centres: Underpinning the Future of Life Sciences and Biotechnology. Retrieved August, 10th from the *OECD* web site: <http://www.oecd.org/>

Stein L. (2002). Creating a bioinformatics nation. *Nature* 417(6885), 119-200

Skerman, V. B. D. (1972). *World Directory of Collections of Cultures of Microorganisms*. New York: John Wiley Sons, Inc.

Sugawara, H, Miyazaki, S., Shimura, J. & Ichiyanagi, Y. (1996). Bioinformatics tools for the study of microbial diversity. *Journal of Industrial Microbiology* 17, pp.490-497

UNESCO (n.d.) Web page of United Nations Educational, Scientific and Cultural Organization for MIRCEN. Available from: http://www.unesco.org/science/life/mircen_network.html/

WDCM (1997) Homepage of WFCC-MIRCEN World Data Centre for Microorganisms. Available from: <http://www.wdcm.org/>

WFCC (n.d.) Homepage of the World Federation Culture Collections. Available from: <http://www.wfcc.info/>

Table 1 The List of searchable databases at <http://www.wdcm.nig.ac.jp/cgi-bin/search.cgi>

CCINFO	The World Directory of Culture Collections: Detailed information on 466 collections in 61 countries is stored.
Bacteria Nomenclature	Bacteria Nomenclature Up-to-date compiled by DSMZ . The search system is developed by WDCM
STRAIN Bacteria	List of 14,000 scientific names in total of bacterial strains preserved in culture collections worldwide
STRAIN Fungi	List of 23,000 scientific names in total of fungal strains preserved in culture collections worldwide
STRAIN Algae	List of algae preserved in culture collections worldwide
STRAIN Archaea	List of achaea preserved in culture collections worldwide
STRAIN cDNA	List of cDNA libraries preserved in culture collections worldwide
STRAIN Cell lines animal	Animal cell lines preserved in culture collections worldwide
STRAIN Cell lines plants	Plant cell lines preserved in culture collections worldwide
STRAIN Hybridomas animal	Hybridoma of animal cells preserved in culture collections worldwide
STRAIN Hybridomas plants	Hybridoma of plant cells preserved in culture collections worldwide
STRAIN Lichens	List of lichens preserved in culture collections worldwide
STRAIN Protozoa	List of protozoa preserved in culture collections worldwide
STRAIN Vectors	List of vectors preserved in culture collections worldwide
STRAIN Viruses Animal	List of animal viruses preserved in culture collections worldwide
STRAIN Viruses Bacteria	List of bacterial viruses preserved in culture collections worldwide
STRAIN Viruses Plants	List of plant viruses preserved in culture collections worldwide
ALGAE	The World Catalogue of Algae Matsushima Mycological Memoirs No.8 Fragmenta Mycologica - I by Keisuke Matsushima and Takashi Matsushima
MMM8	This is an enumeration of novel interesting fungi, where we encountered in mycological works.
Quinone	Search species by quinones
RKCCODE	Coding system of extensive data items (fields) for strain characteristics. Find a code number for your data items by key words.
Catalogues	Catalogues provided by member collections to be installed in the WDCM system