The CABRI website: integrating biological resources information in the bioinformatics network environment

Domenico Marra[1], Federico Malusa[2,3], Francesca Piersigilli[3], Assunta Manniello[1] and Paolo Romano[1]

[1]Istituto Nazionale per la Ricerca sul Cancro di Genova, Largo Rosanna Benzi 10, I-16132 Genova – Italy
{domenico.marra,assunta.manniello,paolo.romano}@istge.it

[2]Sienabiotech Spa, Via Fiorentina 1, I-53100 Siena – Italy
fmalusa@sienabiotech.it

[3]Centro Biotecnologie Avanzate, Larg Rosanna Benzi 10, I-16132 Genova – Italy
f.piersigilli@ebrcn.org

Keywords. Biological Resource Center, database management systems, data integration, bioinformatics

Introduction

Biological resources are essential tools in modern biomedical research [1]. It is therefore essential that information on quality biological resources are well known in the scientific community. Web sites distributing this information are more and more widely available, but access and retrieval of the information through a unique system is highly desirable [2].

The CABRI (Common Access to Biological Resources and Information) project was funded by the European Union (EU) from 1996 to 1999. It aimed at the setting up of a "one-stop-shop" for biological materials and related information. This project led to the setting up of the CABRI web site (http://www.cabri.org/), where catalogues of participating cultures collections could be queried, either individually or collectively, and the Guidelines for the Collection Quality Management that were adopted by partners, could be examined. It includes information on more than 120.000 items from 28 collections including bacteria, filamentous fungi and yeasts strains, human and animal cell lines, plasmids, phages, DNA probes, plant cells and plant viruses from nine centers (BCCM, CABI, CBS, CIP, DSMZ, ECACC, ICLC, NCCB, NCIMB). This wealth of information has been made searchable through an implementation of SRS (Sequence Retrieval Software)[3]. In 2001, a new project was launched, the European Biological Resource Centers Network (EBRCN). This project has been funded by the EU for the period 2001 - 2004. Among its objectives is the extension of the CABRI on-line services, with special emphasis on the achievement of a better integration with molecular biology and literature databanks (see http://www.ebrcn.org/).

Implementation Features

1. Definition of Common Data Sets

CABRI catalogues have been implemented in SRS by first comparing the data structure and contents of collections' internal databases and then defining three distinct data sets for each material. The *Minimum Data Sets* (MDS) consist of the information that is needed to identify a unique item in a catalogue: collection's elements for which this information is not available cannot be inserted into the catalogue, since they are lacking some essential data. The *Recommended Data Sets* (RDS) include supplementary information that is useful in order to achieve an improved description of the characteristics, functions and properties of the material. Although not mandatory, these data should always be included in the catalogue, when available. The *Full Data Sets* (FDS) provide all remaining available information related to the material. CABRI catalogues were independently built and they do not share a common FDS. Each collection has its own.

Proper data input procedures define each field of the MDS and RDS by providing a detailed textual description of its contents and by specifying the input process for the corresponding values. This process can foresee the use of a reference list of agreed values or vocabularies, and/or a predefined syntax for insertion of the data (e.g. for bibliography). Possible applications of a purpose domain ontology is under study.

2. CABRI Simple Search and Shopping Cart

CABRI catalogues can be searched either by the *CABRI Simple Search* interface or by a modified SRS interface, that was adapted from the original one. The former interface offers some advantages: it simplifies the search for those users which are not proficient with SRS and it is able to manage those searches involving synonyms and previous species names in a proper way. The CABRI simple search is able to carry out a two steps, synonym-based search. The searched name is first matched against a reference list of synonyms and alternative names. If it is found, then the right name is

added to the search, so that results will include strains matching both the synonym and the name. The synonym search is implemented by means of original perl scripts. This interface is also particularly useful for a quick retrieval of strains for which either the collection number or the exact name is known. It also allows for a free text search on all the catalogues' contents.

After a fruitful search, returned items can be added to a personal "shopping cart". The shopping cart can be used to temporarily store collection numbers of items that are of interest to the researcher. This information is kept anonymous and can be retrieved during following sessions. Information included in the shopping cart can easily be changed or removed. If desired, a pre-order request of stored items can be sent by e-mail (automatically from the CABRI server) or by fax (after printing locally an ad hoc prefilled request sheet) to involved collections. In this case, personal information is needed for establishing a direct contact between researcher and collections. Personal data are not stored on the web site, but on the personal computer, where it can be kept for later use. The shopping cart is implemented by means of original perl scripts.

3.CABRI HyperCatalogue

In order to improve accessibility of CABRI catalogues through general purpose search engines such as Google, which cannot archive the information included in the SRS system, and to offer a catalogues' navigation tool, the CABRI HyperCatalogue system has been implemented. It is based upon a relational implementation of essential information of all CABRI catalogues and a set of perl scripts and it consists of a set of static web pages which lists all biological resources included in the CABRI catalogues arranged by catalogue, by name and by id (which can be a strain or collection number). These pages are hierarchically structured so that the end user can access the list of items of his/her interest in a few steps only. Detailed information of selected items is finally retrieved through SRS, so that the shopping cart is available when browsing the HyperCatalogue.

4.Links to External Information Sources

An important goal for EBRCN is to add value to current catalogue information and enhance accessibility. This will mainly be implemented by linking catalogue data to bioinformatics resources, such as sequence and genetic databases, and to literature databanks. Links to Medline have already been implemented for some catalogues, while a general procedure to identify and set up links between EMBL Data Library and CABRI catalogues has been defined and will soon start. Further links to external resources have been implemented (eg, plasmids' maps, micro-organisms' images) or being implemented (eg, biochemical pathways' databases [4]).

5.Extracted Databases

The catalogue information will also be offered to other services, so as to maximise the use of the data. Limited, significant information will therefore be periodically extracted from catalogues and made available for downloading to interested public SRS services. This information will be made available both as a flat file and an XML file and adequate SRS configuration files will be distributed together with data,

References

[1] OECD. 2001. Biological Resource Centres, Underpinning the Future of Life Sciences and Biotechnology, Reports of the Organisation for Economic Co-operation and Development, http://www.oecd.org/.

[2] Romano, P., Kracht, M., Manniello, M.A., Stehehuis, G., Fritze, D., Coordinated approaches to the management of biotechnology resources, as it relates to bioinformatics, Applied Bioinformatics (submitted).

[3] Etzold T, Ulyanov A and Argos P. 1996. SRS: information retrieval system for molecular biology data banks. Meth. Enzymol. 266:114-128

 [4] Romano, P., Aresu, O., Manniello, M.A., Parodi, B., Interoperability of CABRI services and biochemical pathways databases, Comparative and Functional Genomics (in press).