# Developments and Perspectives for CABRI Web Services

Daniele Pierpaolo Colobraro, Paolo Romano
IRCCS AOU San Martino IST, Genova, Italy
{danielepierpaolo.colobraro,paolo.romano}@hsanmartino.it

**Abstract.** In this short paper, we introduce the on-going improvements of CABRI Web Services. These are meant to make programmatic access to information on microbial resources much simpler and more effective, by also enabling a real interoperability with some of the most used bioinformatics databases. Such improvements mainly consist in an extension of the returned information, the adoption of the Microbial Common Language (MCL) for data exchange, and of the implementation of REST standard along with the already available SOAP standard.

**Keywords:** microbial biological resources; web services; data interoperability; data integration.

## 1 Introduction

Information provided by Biological Resource Centers (BRCs) is of increasing interest for researchers. Such information consists in catalogues of well conserved and characterized micro-organisms that can be requested from BRCs. These catalogues are made available through BRC web sites, but some portals enables an integrated access to many catalogues, e.g. the Common Access to Biological Resources and Information (CABRI) [1] (http://www.cabri.org/) and the Global Catalogue of Micro-organisms (GCM) [2] (http://gcm.wfcc.info/). CABRI network services offer access to 28 catalogues from European BRCs, including more than 120,000 resources. They were implemented in 1999 as a final deliverable of the CABRI EU project. Catalogues are submitted in a standard format for implementation in a common SRS site.

Much more information exist on these micro-organisms and both CABRI and GCM offer links to external resources, including PubMed, for literature, and the European Nucleotide Archive (ENA), for sequences. Other systems, such as StrainInfo [3] (http://www.straininfo.net/), can offer useful information about strains.

The Microbial Resource Research Infrastructure (MIRRI) is a recently funded European infrastructure, still in its preparatory phase [4] (http://www.mirri.org/). One of its workpackages aims at defining methods for improving the quantity and quality of information on microbial resources, designing a new portal for the distribution of

collections' catalogues, and assessing best ways to improve interoperability with other databases, so that improved software in critical domains, like health and environment, may be developed by leveraging on the future MIRRI information system.

Indeed, a new generation of software able to improve interoperability of biomedical information systems would be useful to support advanced research. Software technologies like Web Services and Workflow Management Systems are being increasingly adopted [5,6,7,8]. To this aim, CABRI Web Services (CABRI-WS) [9] (http://bioinformatics.istge.it/ibws/doc_cabri.html) and the Microbiological Common Language (MCL), an XML based data exchange format for microbiological information [10] (http://www.straininfo.net/projects/mcl) may be essential components for biological resources. CABRI-WS are included in the IST Bioinformatics Web Services (IBWS) that were deployed at the National Cancer Research Institute of Genoa, now IRCCS AOU San Martino IST. IBWS also include SWS (SRS by Web Services), that allow to query databases included in SRS sites, and TP53 Web Services, that retrieve data from an SRS implementation of the IARC TP53 Mutation Database.

Here, we introduce some of the on-going improvements of CABRI Web Services.

## 2 Methods

Distinct WS are available for each biological resource type in CABRI: bacteria and archaea strains, filamentous fungi strains, yeasts strains, plasmids, phages, and human and animal cell lines. Resources can be searched by name, identifier or using free text. Two types of services were implemented: i) searching for name or free text and returning IDs, and ii) searching for an ID and returning full records. The output consists in the contents of CABRI catalogues, formatted as flat files (see Box 1).

```
Strain_number LMG 1(t1)
Other_collection_numbers CCUG 34964; NCIB 12128
Restrictions Biohazard group 1
Organism_type Bacteria
Name Phyllobacterium myrsinacearum, (ex Knösel 1962) Knösel 1984VL
  emend. Mergaert, Cnockaert and Swings 2002 VP
Infrasubspecific_names -
Status -
History <- 1973, D.Knösel (Phyllobacterium rubiacearum)
Conditions_for_growth Medium 1, 25C
Form_of_supply Dried
Isolated_from Pavetta zimmermannia
Geographic_origin Germany, Stuttgart-Hohenheim
Remarks Stable colony type isolated from LMG 1. Type strain of
  Phyllobacterium rubiacearum. See also Agrobacterium sp. LMG 1(t2)
```

**Box 1**. CABRI original flat file format for LMG 1(t1)

CABRI-WS were deployed by using Soaplab, a tool able to provide programmatic access to local, command-line applications and to the contents of ordinary web pages, whose only requirements are Apache Tomcat with the Axis toolkit and a Java Virtual Machine. In Soaplab, new Web Services are added to the system by defining simple

descriptions of related execution commands. These are written in the AJAX Command Definition (ACD) language. CABRI WS can be accessed through any WSDL-SOAP compliant software, including the Taverna Workbench.

CABRI-WS improvements are aimed at: i) extending returned information, ii) adopting MCL for data exchange, iii) implementing REST based Web Services.

We are extending information returned by CABRI-WS by adding data that is available in CCINFO, a directory of culture collections provided by the World Data Center for Microorganism (WDCM), Straininfo, and some reference information systems. MCL is able to represent the contents of CABRI catalogues, with a greater precision and it may be used for the output of the new CABRI-WS. See examples in Box 2. Due to the extension of CABRI-WS data, a revision of MCL may be needed.

**(a)**

```
<mcl:BCR>
   <mcl:WDCMNumber>296</mcl:WDCMNumber>
   <mcl:fullName>Belgian Coordinated Collections of Microorganisms/
LMG Bacteria Colletion</mcl:fullName>
   <mcl:acronym>LMG or BCCM/LMG</mcl:acronym>
</mcl:BCR>
```

**(b)**

```
<mcl:Culture>
   <mcl:strainNumber>LMG 1(t1)</mcl:strainNumber>
   <!-- Strain number from Cabri -->
   <mcl:otherStrainNumber>CCUG 34964</mcl:otherStrainNumber>
   <!-- Strain number from Cabri -->
   <mcl:otherStrainNumber>NCIB 12128</mcl:otherStrainNumber>
   <!-- Strain number from straininfo.net -->
   <mcl:otherStrainNumber>CECT 4452</mcl:otherStrainNumber>
   <!-- Strain number from straininfo.net -->
```

**Box 2**. Information on the BRC from CCINFO (a) and extended information on other strain numbers from Strainfo (b), both excerted from the oputput fomatted with MCL.

| **Web Service** | **Output** | **Call** (prefix http://www.cabri.org/cws/) |
|---|---|---|
| getCataloguesList | List of catalogues | getCataloguesList/views |
| getCatalogue | List of all strains in a given catalogue | getCatalogue/views? name=CABI_BACT |
| getData | Query catalogues | getData/views?resname=bacillus subtilis |

**Table 1.** Summary of REST Web Services

The REST standard has proven to be both effective and simple to adopt. Improved CABRI-WS are being implemented through this standard. To this aim, the WS will be implemented through an Apache web server. Scripts are presently being written in python. In Table 1, a summary of available WS is shown. More information is being made available on-line at http://bioinformatics.istge.it/ibws/rest.html.

## 3 Conclusion

We have presented the main improvements of CABRI WS which are being carried out in the context of the MIRRI project. The new WS include an extended contents, by incorporating data from other information systems. A REST interface is being built and will flank the current SOAP one. The search and retrieval approaches reproduce the features of the standard web interface. The output is based on the MCL language for representation and exchange of microbiological information. These improvements are aimed at simplifying and making interoperability of microbial information more effective, which is one of the main aims of the MIRRI project.

**Acknowledgements.** The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 312251.

## References

1. Romano, P., Kracht, M., Manniello, M.A., Stegehuis, G., Fritze, D.: The role of informatics in the coordinated management of biological resources collections. Appl Bioinf. 4(3):175-186, 2005.
2. Wu L, Sun Q, Sugawara H, Yang S, Zhou Y, McCluskey K, Vasilenko A, Suzuki K-I, Ohkuma M, Lee Y, Robert V, Ingsriswang S, Guissart F, Desmeth P, Ma J. Global catalogue of microorganisms (gcm): a comprehensive database and information retrieval, analysis, and visualization system for microbial resources. BMC Genomics 2013, 14:933.
3. Dawyndt P, Vancanneyt M, De Meyer H, Swings J. Knowledge accumulation and resolution of data inconsistencies during the integration of microbial information sources, IEEE Transactions on Knowledge and Data Engineering, vol.17, no.8, pp. 1111-1126, Aug. 2005
4. Schüngel M, Stackebrandt E, Bizet C, Smith D. MIRRI The Microbial Resource Research Infrastructure: managing resources for the bio-economy. EMBnet.journal 2013, 19(1):5-8.
5. Romano P. Automation of in-silico data analysis processes through workflow management systems, Briefings in Bioinformatics 2008 9(1):57-68.
6. Wolstencroft K, Haines R, Fellows D, Williams A, Withers D, Owen S, Soiland-Reyes S, Dunlop I, Nenadic A, Fisher P, Bhagat J, Belhajjame K, Bacall F, Hardisty A, Nieva de la Hidalga A, Balcazar Vargas MP, Sufi S, Goble C. The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud. Nucleic Acids Research 2013, 41(W1): W557-W561.
7. Bhagat J, Tanoh F, Nzuobontane E, Laurent T, Orlowski J, Roos M, Wolstencroft K, Aleksejevs S, Stevens R, Pettifer S, Lopez R, Goble C. BioCatalogue: a universal catalogue of web services for the life sciences. Nucl. Acids Res. 2010, 38: 689-694.
8. Romano P, Marra D, Milanesi L. Web services and workflow management for biological resources, BMC Bioinformatics 2005, 6(Suppl 4):S24.
9. Zappa A, Miele M, Romano P. IBWS: IST Bioinformatics Web Services. Nucleic Acids Research 2010, 38(Web Server issue):W712-W718.
10. Verslyppe B, Kottman R, De Vos P, De Baets B, Dawyndt P. Microbiological Common Language (MCL): a standard for electronic information exchange in the Microbial Commons. Research in Microbiology 2010, 161(6), 439-445.