

Designing the MIRRI Information System

Culture collections (CCs) manage fundamental data about preserved microorganisms. To enable the exploitation of these resources requires standardized, and simplified, access to the associated information. To this end, a European Microbial Resource Research Infrastructure, MIRRI, was established. A prime objective of MIRRI is to unite and provide universal access to the information available in European public collections of microorganisms through a dynamic Information System, the MIRRI-IS. In this context, the desired outcome would include a repository for microbial domain Biological Resource Centre (mBRC) catalogues in a common format, the validation of catalogue contents to provide quality data, the interconnection of domain information systems for data extension, a unique portal for catalogues and associated data and an interoperable system based on Application Programming Interfaces (APIs) and service based interfaces and workflows.

The architectural design of the system foresees the adoption of a standard format for exchanging data between CCs, a Minimum Data Set for essential data, to evolve into Minimum Information about Biological Resources (MIaBRe), a user-friendly interface, as well as APIs and services/workflows for integration software.

Three systems were developed in the MIRRI preparatory phase. They have been developed to cope with distinct issues. The BacDive demonstrator aims at extending the contents of catalogues with a greater number of better defined data. The StrainInfo demonstrator is targeted towards a better integration among collection catalogues through the identification of common strains. It makes some order in strains available in various collections and makes possible the re-organization of collections and the sharing of data between catalogues. The USMI Galaxy demonstrator is aimed at supporting data curation and at integrating catalogues with external resources. It makes it possible to integrate collections' data with other bioinformatics databases by leveraging on existing tools and with little development requirements. Moreover, it allows the improvement of a collection's data by automating links to external databases.

A five-year plan for the implementation of the MIRRI-IS has been defined. The four main lines included in the plan are data curation, interoperability, applications development and IT competence. Data curation is meant to be developed by the progressive adoption of Standard Operating Procedures, able to lead user catalogues from the current Minimum Data Set to a structured extended data set and finally to data sets fully compliant with Minimum Information about Biological Resources (MIaBRE) guidelines, still under definition. Interoperability will lead current separated catalogues to interact with an integrated repository of reference data which will finally allow user access to all catalogues through a common portal.

Designing the MIRRI Information System

P. Romano¹, D. Smith², B. Bunk³, A. Vasilenko⁴, F. O. Glöckner⁵

¹ IRCCS AOU San Martino IST, Genoa, Italy

² CABI, Egham, United Kingdom

³ Leibniz-Institut DSMZ, Braunschweig, Germany

⁴ All-Russian Collection of Microorganisms (VKM), Pushchino, Russia

⁵ Jacobs University Bremen, Bremen, Germany

Introduction

Culture collections (CCs) manage fundamental data about preserved microorganisms. However, information currently accessible in their databases rarely adheres to common standards. The current status sees distinct on-line catalogues with limited and heterogeneous content, data models and formats, different user interfaces, very little inter-connections and only a few links to external databases. To enable the exploitation of these resources requires standardized, and simplified, access to the associated information. To this end, a European Microbial Resource Research Infrastructure, MIRRI, was established (1). A prime objective of MIRRI is to unite and provide universal access to the information available in European public collections of microorganisms through a dynamic Information System, the MIRRI-IS. In this context, the desired outcome would include a repository for microbial domain Biological Resource Centre (mBRC) catalogues in a common format, the validation of catalogue contents to provide quality data, the interconnection of domain information systems for data extension, a unique portal for catalogues and associated data and an interoperable system based on Application Programming Interfaces (APIs) and service based interfaces and workflows.

Methods

The architectural design of the system (figure 1) foresees the adoption of the following components: a standard format for exchanging data between CCs likely to be based on the Microbiological Common Language (MCL), a Minimum Data Set for essential data, to evolve into Minimum Information about Biological Resources (MIaBRe), a user-friendly interface to be included in a Collaborative Working Environment (CWE), as well as state-of-the-art APIs and services/workflows for well-known and largely adopted integration software.

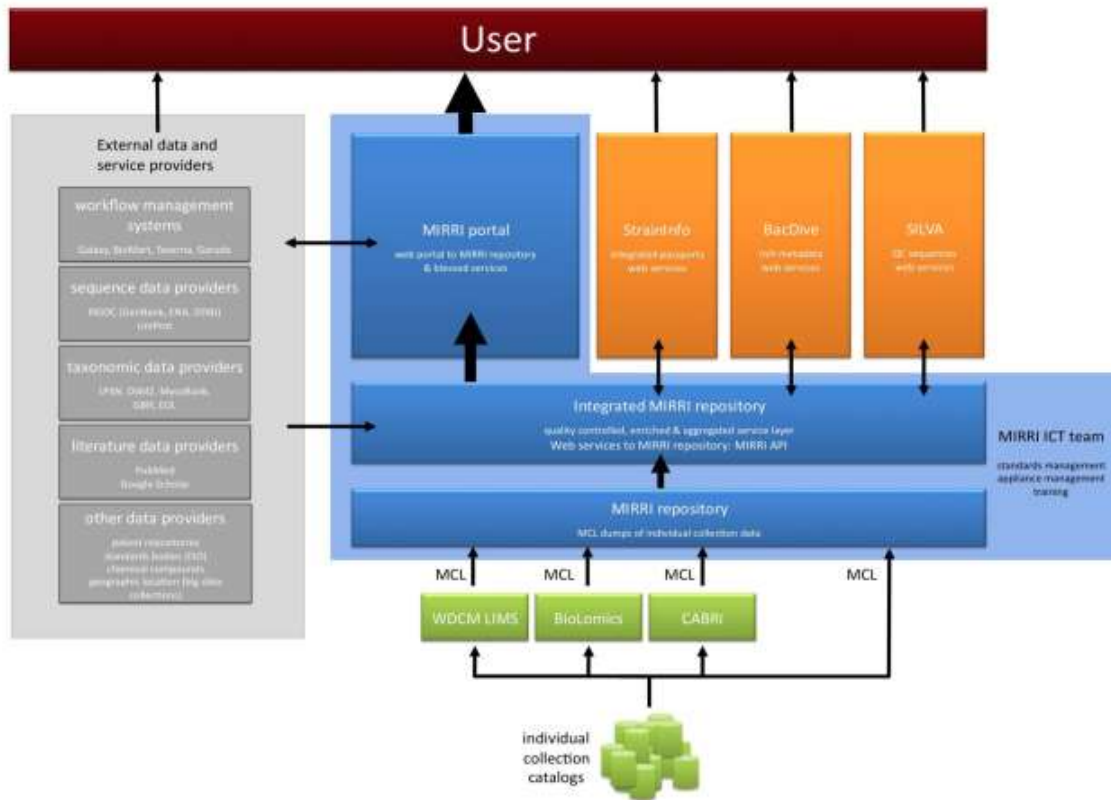
Results

On the way to achieve the desired outcome, three systems, the so called "MIRRI demonstrators", were developed in the MIRRI preparatory phase. The three demonstrators have been developed to cope with distinct issues, all equally essential towards the creation of a MIRRI Information System. The BacDive demonstrator aims at extending the contents of catalogues with a greater number of better defined data. The effort put into the building of BacDive is greater than what can usually be done by a generic culture collection. However, it demonstrates which information could be useful (for a given organism type) and designs a way to manage all this information. Then, it shows how this "content extension" can be achieved progressively, by selecting subdomains of interest starting from the most recent/interesting strains.

The StrainInfo demonstrator is targeted towards a better integration among collection catalogues through the identification of common strains. It makes some order in strains available in various collections and makes possible the re-organization of collections and the sharing of data between catalogues.

The USMI Galaxy demonstrator is aimed at supporting data curation and at integrating catalogues with external resources. It makes it possible to integrate collections' data with other bioinformatics databases by leveraging on existing tools, like Galaxy, well known and with little development requirements. Moreover, it allows the improvement of a collection's data by automating links to

external databases that may help the adoption of standardized, and shareable, terminologies and data values.



Conclusions

A five-year plan for the implementation of the MIRRI-IS has been defined. The four main lines included in the plan are data curation, interoperability, applications development and IT competence. Data curation is meant to be developed by the progressive adoption of Standard Operating Procedures, able to lead user catalogues from the current Minimum Data Set to a structured extended data set and finally to data sets fully compliant with Minimum Information about Biological Resources (MIaBRE) guidelines, still under definition. Interoperability will lead current separated catalogues to interact with an integrated repository of reference data which will finally allow user access to all catalogues through a common portal.

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. MIRRI: <http://www.mirri.org/>
2. Bunk B, Colobraro DP, Dawyndt P, De Vos P, Glöckner FO, Kopf A, Robert V, Romano P, Smith D, Söhngen C, Van Hauwenhuysse F, Vasilenko A. D8.4_MIRRI_Report on comparison of various integration software operating in Life Sciences. <https://zenodo.org/record/49968>
3. BacDive: <http://bacdive.dsmz.de/>
4. StrainInfo: <http://straininfo.net/>
5. USMI Galaxy: <http://bioinformatics.hsanmartino.it:8080/>