

SyMBA: A Generic Data and Metadata Archive for Multi-Omics Experiments



Allyson Lister, Frank Gibson,
Matthew Pocock, Anil Wipat
CISBAN, Newcastle University



symba-devel@lists.sourceforge.net • <http://symba.sourceforge.net>

Integrated Information

SyMBA is an archive that facilitates metadata integration and data storage. It stores knowledge about experimental data in a computationally-amenable and user-friendly way.

It is designed to prevent deletion, loss, or accidental modification of primary data and metadata, while providing convenient manual and computational access for standards-compliant publication, sharing and analysis.

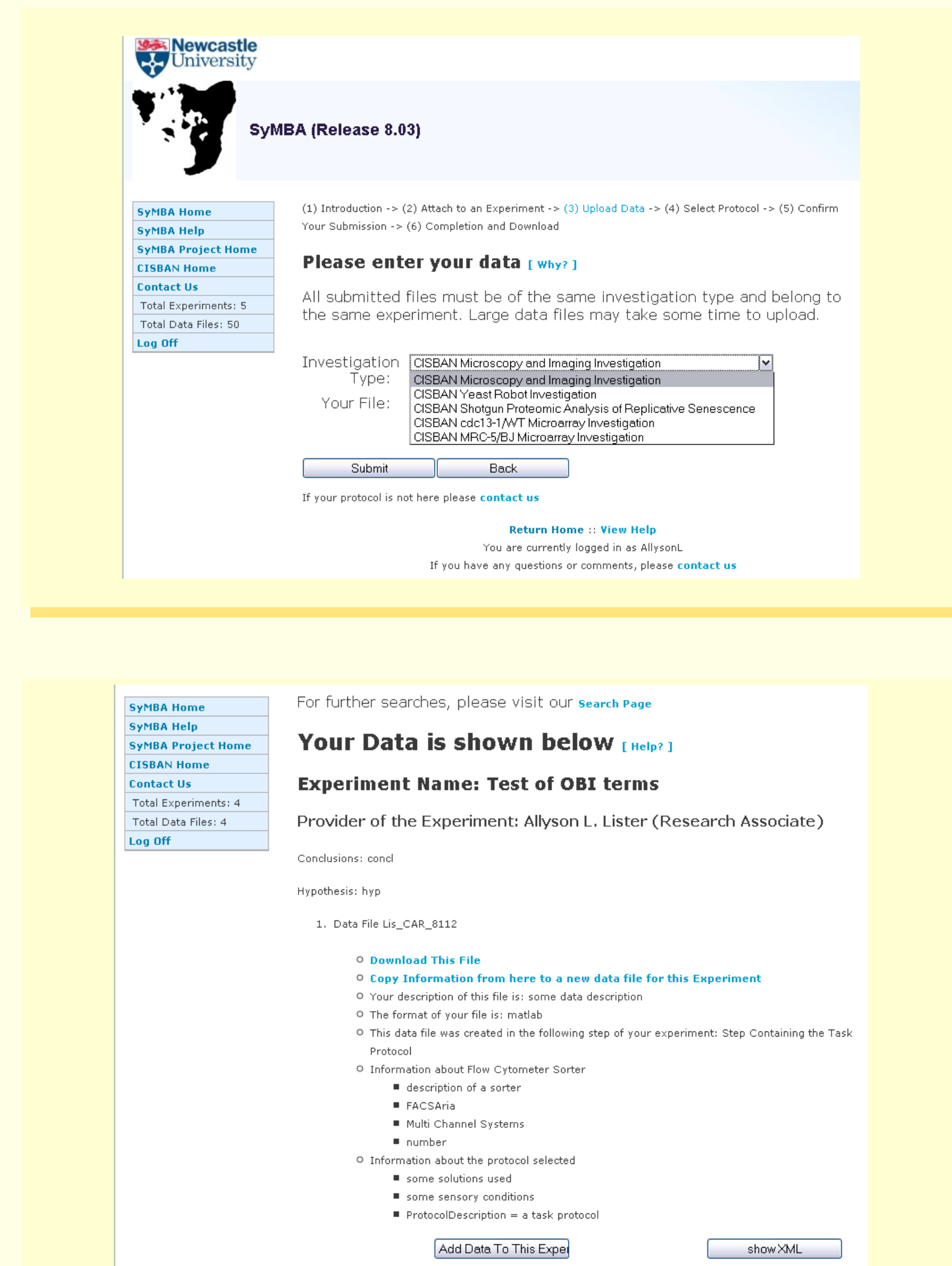
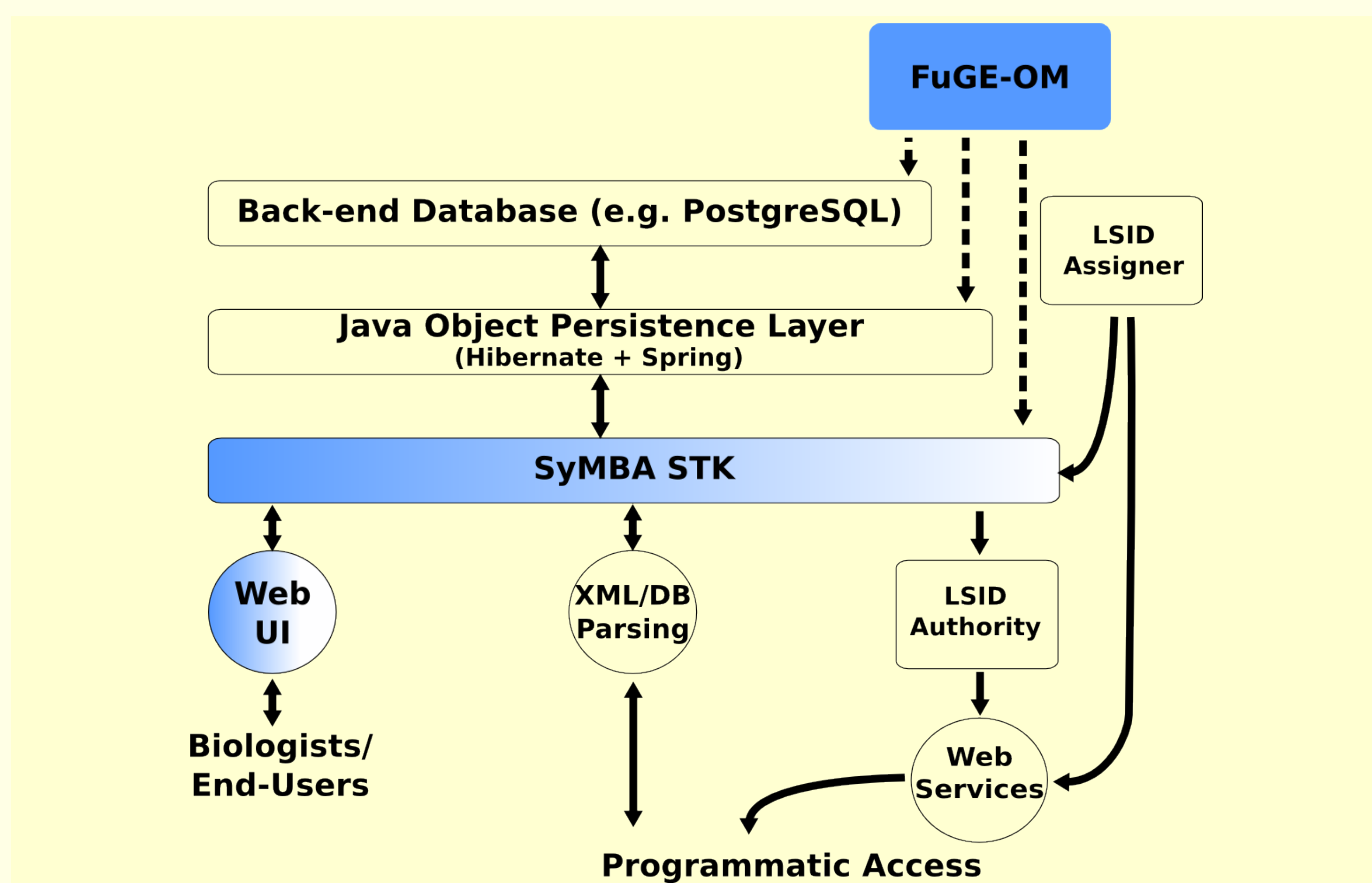
Metadata is stored in a single format, irrespective of the type of data it describes. SyMBA presents users with a simple, unified view of their data and metadata, while utilizing the richly-descriptive Functional Genomics Experiment Object Model (FuGE-OM[1], <http://fuge.sf.net>) behind the scenes.

Customizable

- SyMBA can be customized for specific needs of users and developers
- Any type of life-science experiment may be modelled within SyMBA
- Each installation of SyMBA includes a metadata repository and data store that can be customised to meet specific needs
- The basic protocol for each new experiment is stored by the SyMBA installer as a “template” in the SyMBA database
- These templates are used to customize the dynamically-generated data entry forms within the SyMBA website
- Templates can be built to community standards such as MIAME[2] and MIAPE[3]

SyMBA Inside and Out

On the left is an overview of the SyMBA architecture. Dark shading signifies hand-written code, while unshaded areas designate automatically-generated code. Circles mark agent access points. Solid arrows indicate direction of data flow. The LSID Authority, which returns data associated with a given LSID, and the LSID Assigner, which creates new LSIDs, are accessible via web services. The other two images are screenshots of viewing an experiment and creating a new experiment.



Standards-Driven

CISBAN is a participant in the RSBI working group within MGED, which promotes the use of reporting standards for biological investigations in three ways:

- **FuGE**, <http://fuge.sf.net>, provides a standard format for experimental metadata by modelling the common components of scientific experimentation such as samples, protocols, instruments and software
- **MIBBI**, <http://mibbi.sf.net>, a catalog of checklists such as MIAME
- **OBI**, <http://purl.obofoundry.org/obo/obi>, an integrated ontology for the description of biological and clinical investigations

SyMBA uses FuGE as the core of its database structure, and provides examples of use for both MIBBI checklists and OBI terms.

What You Get

- A browsable view of experimental data and metadata
- The ability to compare aspects of studies such as controls and treatments
- A unified and simple web interface for data capture
- Programmatic access via web services and the STK
- Improved data organisation via the use of FuGE
- The source code is available under the LGPL from <http://symba.sourceforge.net>
- For evaluation of the SyMBA web interface, a test installation is available at <http://www.cisban.ac.uk/symba-sandbox> (Username: MickeyM, Password: MousePass)