SIMBioMS: services for collaborative projects.

Fernandez-Banet J.[1], Perheentupa T.[2], Heikkinen J.[2], Viskna J.[3], Gostev M.[1],Shen Huei-Yi[2], Brazma A.[1], Krestyaninova M.[1], and ENGAGE consortium.
[1] European Bioinformatics Institute, Cambridge, UK.
[2] Institute for Molecular Medicine, University of Helsinki, Helsinki, Finland.
[3] Institute of Mathematics and Computer Science, University of Latvia, Riga, Latvia.

The current trend on scientific studies to move to high impact, complex collaborative studies and the amount of information generated by those, have increased the demand on systems to aid in the management and secure exchange of data.

Currently most biological databases are designed as long-term data repositories with very strict requirements that not always suit the needs of specific scientific projects.

SIMBIOMS (http://ww.simbioms.org), a modular information management system, has been developed as an open source solution for collaborative projects in biology that need a highly customizable, easy to configure solution for efficient data sharing and storage. Due to the multi-partner nature of many new projects where different groups generate/use different data, SIMBIOMS also introduces tools to aid in the management and administration of data access. SIMBIOMS is comprised of four modules that can be used independently or integrated as a single system.

These modules functionality is describe below.

AIMS (Assay Information Management System), acts as a central repository for experimental data produced by different technologies (genomic, transcriptomic, proteomic, etc). Data is organized in three categories being the most common set up: Study, Experiment, and Assay. Each category can store metadata specific for that level as well as accompanying data (experiment input/output data) and supplementary data files.

Among AIMS functionalities there are tools for metadata report generation, concurrent export of multiple data files and tools to facilitate the upload of results data into long-term repositories as requested for publication.

SIMS (Sample Information Management System) is a system to allow researchers to track information about samples, such as date of collection or storage conditions as well as phenotype data related to the Samples. As with AIMS, it presents data at three different levels, being a most common configuration: Person (first level), Sample (second level), and Aliquot (third level). The system is designed in such a way that only anonymous data is allowed to ensure data confidentiality.

AIMS and SIMS can work as a united system where AIMS holds data and metadata about the experiments/assays while SIMS holds metadata about phenotype and sample conditions.

SAIL is a solution designed to address the need to browse, index and annotate the exponentially increasing amount of data generated by biomedical studies. The system is designed to hold phenotype availability information and meta-data about samples and experiments, submitted by data owners or databases. SAIL serves as an index of phenotype availability in different collections which can be used in different scenarios, from the design of new studies to finding biomaterial from different cohorts that can be combined in meta-analysis studies.

RepTool is a tool to ease the management and resource allocation and the generation of reports on the state of new or existing studies. This tool integrates with AIMS and SIMS allowing users to manage their accounts (change passwords, etc) and eventually open and close studies for submission.

On the production side and as a service provider, SIMBIOMS can also act at different levels, from assisting in the definition of project requirements (which data needs to be shared and how it should be structured) to hosting and rapid system deployment (guarantying data security and integrity).

SIMBIOMS has been successfully implemented and used in high impact collaborative projects like ENGAGE (www.euengage.org) and MOLPAGE (www.molpage.org).