# SIMBioMS: open source system for information management on 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 latest developments in scientific research technology have paved the way to a move to more complex, high impact collaborative studies where the vast amount of data generated and the need to share that data have increased the demand on systems to assist the management and secure exchange of information.

While most biological databases act as long term data repositories with, in many cases, very strict rules on what type of data can be allowed and how it should be formatted, the offer of more flexible data management systems oriented for short or medium term storage that would better suit the needs of specific scientific projects is still weak.

SIMBioMS (http://ww.simbioms.org), an open source modular information management system, has been developed as an easy to use, very flexible solution for collaborative projects in biology with an urge for efficient data sharing and storage.

Due to the nature of multi-partner projects different types of data may be accessible for a restricted number of users, while data generation and analysis may also require different sets of user access rights. For this matter, SIMBioMS also introduces tools to aid in the management of user rights and administration of data access.

Current SIMBioMS services are based on four open soruce software modules and these can be used separately on in combination with each other.

AIMS (Assay Information Management System) and SIMS (Sample Information Management System) are the two main modules. They consist in a database with an associated file system and a web interface that act as the central repositories of data and metadata. Both systems share a similar design where data is structured hierarchically in three different levels where each level can be customized to hold any relevant metadata.

AIMS focus is on experimental data produced by different omics technologies. While the structure of the data is flexible and the system can be easily configured to suit other set ups, the most commonly used one is: Study, Experiment, and Assay.

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 role would be tracking information about samples, such as date of collection or storage conditions as well as phenotype data related to the Samples. The most common set up for SIMS would be: Person (first level), Sample (second level), and Aliquot (third level).

Due to the sensitive nature of phenotype data, SIMS has been designed in such a way that only anonymized data is allowed to ensure data confidentiality.

AIMS and SIMS can work as a united system where entries in both systems can be linked enabling the possibility of generating complex metadata reports.

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.

Emanta 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, SIMBioMS can 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 (sftp sandbox and firewall set up along side automatic data backup guarantying data security and integrity), also offering a unified structure for services support and maintenance.

SIMBioMS has been successfully implemented and used in projects like ENGAGE (www.euengage.org) and MOLPAGE (www.molpage.org).