

## Research projects for MSc theses

### Consolidation of a semantic data repository for microbial biofilms

#### Supervisor

Dr Andreas Bohn  
mail: abohn@itqb.unl.pt

#### Laboratory

Systems Biodynamics Group, ITQB-UNL, Oeiras, Portugal  
<http://sbd.itqb.unl.pt>

#### Area of research

Computational Biology / Ecoinformatics  
Microbial ecology

#### Project summary

Phototrophic biofilms are microbial consortia attached to illuminated submerged surfaces, consisting of phototrophic cyanobacteria, diatoms, and algae, which are encased in symbiosis with heterotrophic bacteria in a matrix of extracellular polymeric substances.

Besides being an important object of study for the understanding of light-driven ecosystems, these biofilms have potential applications in biotechnologies like aquafarming or wastewater treatment.

In the frame of a former European project on phototrophic biofilms (PHOBIA, <http://www.photobiofilms.org>), a data collection englobing a large array of variables representing the species composition, the spatial and temporal structure, and the metabolism of these microbial communities was created. The management of this information and the corresponding knowledge discovery is a typical task for the recent field of ecological informatics, dealing with the challenges emerging in the treatment and the analysis of highly heterogeneous datasets, comprising multi-scale data from genes to the biosphere.

Organizing such highly heterogeneous data with relational models is prone to creating very complex database structures, which are difficult to modify and are in extreme cases only understood by the expert who created the database, thus hampering the reusability and persistence of the database. Computational approaches from the semantic web have been appointed as a solution to circumvent this problem and improve the interoperability of data repositories.

To implement a database for the PHOBIA data, the Systems Biodynamics Group has used the S3DB software (<http://www.s3db.org>), a computational framework consisting of a logical data model built with semantic web tools, and a module for public and private access control to the data. Applying some modifications to the original software, we could tailor it for our purposes and this way improve the performance of the database.

The objective of the present project is to compare our S3DB approach with other current softwares for semantic or NoSQL databases, which are presently being developed at a fast pace, and to improve and eventually substitute the current PHOBIA database with the most efficient approach available.

The workplan thus consists in 1) a literature research on current approaches to semantically aware data management, 2) comparison and benchmarking with the S3DB software, 3) implementation of the most efficient approach emerging from the comparison, and 4) optimization of the underlying data model and user interface.