

Designing an Ontology Tool for the Unification of Biofilms Data

Ana Margarida Sousa¹, Maria Olívia Pereira¹,
Nuno F. Azevedo², and Anália Lourenço^{1,3}

¹ CEB - Centre of Biological Engineering, University of Minho,
Campus de Gualtar, 4710-057 Braga, Portugal
{anamargaridasousa, mopereira}@deb.uminho.pt

² LEPABE – Dep. of Chemical Engineering, Faculty of Engineering, University of Porto,
Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal
nazevedo@fe.up.pt

³ ESEI - Escuela Superior de Ingeniería Informática, Edificio Politécnico, Campus
Universitario As Lagoas s/n, Universidad de Vigo, 32004 Ourense, Spain
analialia@ceb.uminho.pt, analialia@uvigo.es

Abstract. The description of biofilm features presents a conceptual and practical challenge. Biofilm studies often encompass multidisciplinary approaches from Biology, Chemistry, Medicine, Material Science and Engineering, among other fields. Standardising biofilm data is essential to be able to accomplish large-scale collaborative and complementary analysis. To define a common standard format to exchange the heterogeneous biofilm data, it is first necessary to define a set of minimum information for the documentation of biofilm experiments. Then, data should be organised and semantically integrated. This paper describes the first ontology designed to share structured vocabulary for the annotation of the general biofilm experimental workflow – the Biofilm Ontology (BO). This ontology is intended for the broad research community, including bench microbiologists, clinical researchers, clinicians, curators and bioinformaticians.

Keywords: Biofilms, minimum information about a biofilm experiment, ontology, functional modelling.

1 Introduction

Biofilms are organised communities of microorganisms attached to each other and/or to a surface, and involved in a self-produced polymeric matrix [1]. Biofilms are ubiquitous to natural, clinical and industrial environments and thus, their ecological impact is transversal to many economic and social areas [2].

The study of biofilms is a multidisciplinary knowledge field, at the crossroads of Biology, Chemistry, Medicine, Material Science and Engineering, among others. Biofilms are extremely complex environments and their function resembles that of a multicellular organism [1]. Depending on the biofilm ecological niche, and the particular relationships established among species, microorganisms have different

metabolic and genetic profiles [3]. Notably, each biofilm presents a unique biological signature. Therefore, biofilm studies often encompass multidisciplinary approaches to the characterisation of the structure and activity of the sessile community. For example, these studies may include mass spectrometry, microscopy, flow cytometry, antimicrobial susceptibility, respiratory activity, metabolomic, proteomic, transcriptomic and genomic techniques. Consequently, their outputs tend to vary greatly in type and semantics, including numeric, image, and spectra data (Table 1).

Table 1. Characteristic types of biofilm data and techniques used

Type of biofilm data	Technique
Numeric	Spectrophotometric methods (CV, XTT, ATP detection, Lowry protein assay, Dubois assay and Alamar blue), CFU, antimicrobial susceptibility
Images	Microscopy techniques (SEM, TEM, CLSM, FISH), colony morphology characterisation, gram-staining, proteomic techniques (SDS-PAGE, 2D electrophoresis)
Spectra	Mass spectrometry, MALDI-TOF, chromatography

Legend: CV – crystal violet, XTT - 2,3-Bis-(2-Methoxy-4-Nitro-5-Sulfophenyl)-2H-Tetrazolium-5-Carboxanilide; CFU – colony-forming units; SEM – scanning electron microscopy; TEM – transmission electron microscopy, CLSM – confocal microscopy, FISH - fluorescence *in situ* hybridization, SDS-PAGE - sodium dodecyl sulfate polyacrylamide gel electrophoresis; MALDI-TOF - Matrix Assisted Laser Desorption/Ionization Time of Flight.

Currently, the Minimum Information About a Biofilm Experiment (MIABiE) international consortium (<http://miabie.org>) is working on the definition of guidelines to document biofilm experiments and the standardisation of the nomenclature in use. Specifically, this consortium has established the minimum set of information to be recorded and published about an experiment in order for the procedure and results to be unambiguously and comprehensively interpreted [4]. These issues were discussed with field experts - at the Eurobiofilms 2013 meeting in Ghent, Belgium (9-12 September 2013) - as means to accommodate for as much of this complexity and variability as possible, and anticipate new requirements resulting from the increasing use of high-throughput methods. The creation of controlled vocabulary in support of the description of biofilm studies is inherent to this standardisation effort. Controlled vocabulary based on well-engineered ontologies may support powerful querying and computational analysis tools. Therefore, this paper brings forward the development of the first ontology on Biofilms, identifying the main areas of terminology to be accounted for as well as discussing its integration with other upper and domain ontologies, such as ontology for scientific experiments (EXPO), Chemical Entities of Biological Interest (ChEBI), and Functional Genomics Ontology (FuGO). The adequateness and extensibility of the new ontology, named Biofilm Ontology (BO), has been discussed with MIABiE members and other domain experts, and validated against different studies in the BioOmics, the centralised database on biofilms experiments [5]. Here, we introduce BO and exemplify some of its main descriptive abilities.

2 The Organising Principles of the Biofilm Ontology

There is no standardised methodology for building ontologies. However, Open Biological and Biomedical Ontologies (OBO) foundry has introduced some useful guidelines and principles regarding the different stages of the ontology development life-cycle that helped the construction of the BO [6]. In particular, the organisation of the BO was based on the following main criteria:

- BO is restricted to the biofilm knowledge domain and, therefore, it contains just model concepts and relations that are relevant to the representation of biofilm data;
- BO should be used for annotating data in databases and for textual documentation as such, it should be understandable to people and unambiguously interpreted by software;
- BO development should be pragmatic, that is as new devices, techniques or applications arise, it should be possible to integrate new branches without affecting the existent ontology structure;
- any biofilm experiment should be comprehensively described by a combination of BO instances;
- whenever possible, terms should have a synonyms list to avoid misinterpretations and to enable consistent data curation and repositories searching;

Then, BO development followed the typical ontology life-cycle previously described [7]. The purpose and scope of the ontology were well identified and centred on the Biofilm domain, following MIABiE directives. Knowledge acquisition relied on varied sources of information, namely: discussion of the BO with other Biofilm experts, metadata associated to the experiments in BioOmics database [5], research papers on biofilms, and some learning from other bio-ontologies. The construction of the ontology undertook the conceptualisation and integration of the biofilms concepts identified, formally defining their properties, meaning and relationships with other concepts.

Top-down and bottom-up approaches were combined. First, the top-down approach led to the insertion of the BO into a generic upper ontology. Generic ontologies describe general and domain-independent knowledge, aiming to avoid the duplication of terms related to template structures typical of scientific experiments of any research field. As such, BO would focus only on the specificities of biofilms studies, delegating general experimental characterisation to upper ontologies. Then, the bottom-up approach was implemented to gather and organise the biofilms-specific concepts into ontological instances and establish relations among them. Since Biofilms are a multidisciplinary knowledge field, many concepts will be cross-linked to ontologies in related domains, avoiding term duplication and enforcing data interoperation across platforms and resources, in benefit of broad research community.

BO was represented in formal language and comprehensively documented to promote its use in databases, bioinformatics tools and other resources created in

support of Biofilm research. Its consistency, completeness and conciseness was evaluated by practical exercise, i.e. applying its terms to the description of published studies, and by discussing the ontology with field experts at international conferences, such as the Eurobiofilms 2013 meeting in Ghent, Belgium (9-12 September 2013).

3 The Structure of the Biofilm Ontology

BO development follows the principles of the Open Biological and Biomedical Ontologies (OBO) foundry [6]. Terms are organised in a hierarchical structure with four main branches: biofilm model, biofilm growth conditions, biofilm removal and biofilm characterisation (Fig 1). These branches represent the main steps/components of a typical biofilms experiment. Next, we will provide a brief overview of each of these branches.

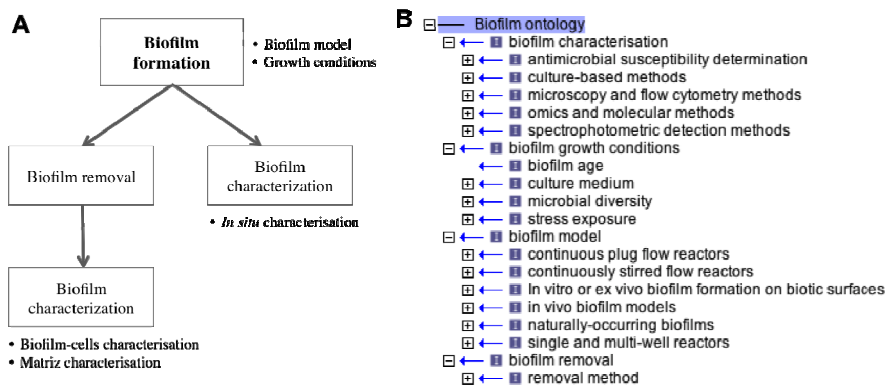


Fig. 1. (A) Typical biofilm experiment workflow; (B) Hierarchical structure of the BO. The four main branches, which compose BO, ‘biofilm characterisation’, ‘biofilm growth conditions’, ‘biofilm model’, and ‘biofilm removal’ represents the typical steps of a biofilm experiment. Each branch is composed by multiple child-terms that, by its turns, have also their descendent terms as denoted by the signal ‘+’. All the BO terms are related among them through the “is_a” relationship type.

Biofilm Model

This BO branch intends to encompass all systems in which biofilms can be formed, analysed or retrieved, including the models of naturally-occurring biofilms, single and multi-well reactors, continuously stirred flow reactors, continuous plug flow reactors, *in vivo* biofilms, and *in vitro* or *ex vivo* cultivation of biofilms on biotic surfaces [8].

Biofilm Growth Conditions

Besides identifying the model used to form the biofilms, the description of the experiment must also specify the operational conditions. These conditions include temperature, culture medium and supplements, hydrodynamic conditions (static or dynamic conditions), pH, oxygen availability (aerobic, anaerobic or microaerophilic conditions), time of biofilm growth and maturation, and stress conditions.

Biofilm Removal

Typically, the characterisation of biofilm cells requires the removal of the biofilms from surfaces where they were formed and further cell detachment. This is a critical step that may alter the physiological state of cells and bias the results of the experiment. This module of the BO collects the methods or techniques commonly used, such as sonication or scrapping, and the operation conditions, such as the device used, period of time of biofilm removal (for example, 5 cycles of 30 s of sonication) and solvent used to collect biofilm-cells (for example, water, phosphate buffer or other buffer solution).

Biofilm Characterisation

This branch describes the methods according to the type of analysis conducted, as follows: culture-based methods, such as CFU counting and colony morphology characterisation; non culture-based methods, such as DAPI, CV and FISH; microscopy (SEM, TEM and CLSM) and flow cytometry methods; spectrophotometric methods, such as XTT, ATP detection and Lowry protein assay; “Omics” and molecular methods, such as proteomics, genomics; and antimicrobial susceptibility testing, such as determination of the minimum inhibitory and bactericidal concentrations and the minimum biofilm eradication concentration. Techniques that are common to other research areas are included, but BO delegates on their definition into well-established domain-specific ontologies and controlled vocabularies (Table 2).

Table 2. Ontologies or controlled vocabularies (CV) relevant in biofilm field and cross-reference by BO

Ontology/ controlled vocabulary	Source
Chemical entities of biological Interest (CHEBI)	http://obo.cvs.sourceforge.net/obo/obo/ontology/chemical/chebi.obo
Colony morphology characterization (CMO)	http://mibbi.sourceforge.net/projects/MIABIE.shtml
Gene ontology (GO)	http://obo.cvs.sourceforge.net/obo/obo/ontology/genomic-proteomic/gene_ontology.obo
Functional Genomics Investigation Ontology (FuGO)	http://sourceforge.net/projects/fugo/
MALDI imaging ontology (IMS)	http://www.maldi-msi.org/download/imzml/imagingMS.obo
PSI-Mass Spectrometry CV (MS)	http://psidev.cvs.sourceforge.net/viewvc/psidev/psi/psi-ms/mzML/controlledVocabulary/psi-ms.obo
PRIDE CV	http://code.google.com/p/ebi-pride/source/browse/trunk/pride-core/schema/pride_cv.obo
Protein ontology (PRO)	http://obo.cvs.sourceforge.net/obo/obo/ontology/genomic-proteomic/pro.obo
PSI-Sample Processing and Separations (SEP)	https://psidev.svn.sourceforge.net/svnroot/psidev/psi/sepcv/trunk/sep.obo

4 Cooperation and Integration with Other Bio-Ontologies

There are several upper ontologies that may assist BO in describing generic experimental setup and procedures. None of these ontologies is an ideal representation of general knowledge and, therefore, there must be a compromise between the “imperfection” degree of the upper ontology and the practical needs of BO, as practical domain ontology. Here, the Suggested Upper Merged Ontology (SUMO) [available at <http://www.ontologyportal.org/>], proposed by the IEEE Standard Upper Ontology Working Group, was selected to formalise concepts that are seen as meta, generic or abstract to a broad range of domain areas (e.g. medical, financial, engineering).

BO was also integrated with the ontology for scientific experiments (EXPO), which includes the fundamental concepts about experiment design, methodology, and the representation of results that are domain independent [9]. For example, any experiment has a goal (‘EXPO:ExperimentalGoal’) and it aims to test a hypothesis (‘EXPO:ExperimentalHypothesis’) that results may confirm it (‘EXPO:FactSupport’) or reject it (‘EXPO:RejectSupport’).

BO integration and cooperation with other ontologies is not limited to generic or top-level ontologies. BO also cooperates with other domain ontologies. BO is focused on biofilm-specific data issues and, therefore, data coming from other knowledge domains, such as flow cytometry, proteomic techniques or microarrays, should be annotated according to the data standards of the respective consortia (Table 2).

5 Applications of BO

The availability of BO allows consistent documentation of biofilm experiments, and facilitates large-scale and computer-aided data processing and analysis. Ultimately, the BO aims to support the comparison of inter-laboratory experiments and the generation of new experimental hypotheses.

BiofOmics [5], the first ever public repository dedicated to biofilm experiments, is a biological resource that takes advantage of this ontology. At first, BO will assist in intelligent experiment screening and comparison, recognising experiments with similar setup and goals of analysis. By using the BO, researchers may compare their own data against other datasets in the repository. Specifically, by means of ontology-based meta-analysis it is possible to compute and score correlations between datasets and a given ontology term, e.g. “biofilm susceptibility”, “matrix composition”, “viable cells”, or “biofilm mass”. This sort of analysis across published biofilm data may provide insights into important research questions, such as: 1) the identification of consistently expressed genes by microorganisms forming biofilms; 2) the description of the mechanisms of resistance and persistence of biofilms in numerous circumstances, e.g. resistance to antibiotic treatments and to the host immune defences, persistence to environmental stress, including pH, microbial predation, or starvation; 3) the identification of microbial biomarkers that may guide the development of new drug therapeutics. In addition, the BO represents a valuable

resource to data curation. With its support, curators are able to interpret and validate richer and more detailed experimental setups and results, assisting data submitters in the comprehensive description of their data. The utility of BO is better demonstrated with some examples of descriptions of biofilm experiments extracted from literature (Fig. 2). Typically, studies including biofilm experiments describe the conditions in which biofilms are formed and the subsequent analysis performed to characterise the community under study.

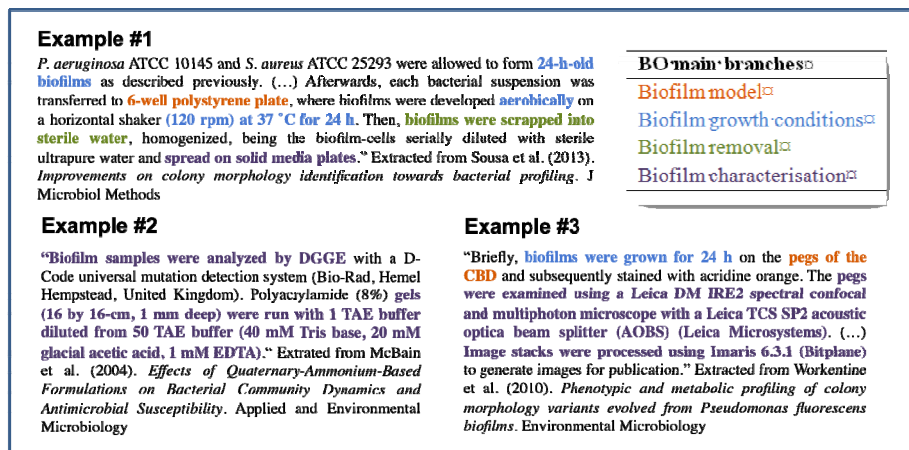


Fig. 2. Examples of the use of BO terms in the description of biofilms studies

6 Conclusions

Biofilms remains a major challenge to several scientific fields. The global context of biofilms makes data integration essential for sharing and retrieval the scientific information that semantic annotation using to ontologies have a crucial role. Despite there are several ontologies of specific knowledge domains, it was verified a lack of semantic resources for biofilm-specific terms. This paper describes the development of the Biofilm Ontology (BO), which aims to address the issue of representing biofilm information, achieving interoperability between scientific areas, research groups and laboratories.

In this paper, it was presented the current state of BO and how it is related to other ontologies, upper and domain ontologies, relevant for the biofilm domain. The actual version covers all steps of a typical biofilm experiment workflow, including biofilm formation (composed by the branches 'biofilm model' and 'biofilm growth conditions'), 'biofilm removal' and 'biofilm characterisation'. However, BO is prepared for future alterations. BO design contemplates existing procedures and dependencies, but it is flexible to account for future extensions. Through active dissemination and group discussion, the biofilm community is invited to collaborate in the population and update of BO.

Acknowledgements. The authors thank the projects PTDC/SAU-ESA/646091/2006/FCOMP-01-0124-FEDER-007480FCT; PEst-OE/EQB/LA0023/2013; “BioHealth - Biotechnology and Bioengineering approaches to improve health quality”, NORTE-07-0124-FEDER-000027, co-funded by the Programa Operacional Regional do Norte (ON.2 – O Novo Norte), QREN, FEDER; RECI/BBB-EBI/0179/2012 - Consolidating Research Expertise and Resources on Cellular and Molecular Biotechnology at CEB/IBB, FCOMP-01-0124-FEDER-027462, FEDER; and the Agrupamento INBIOMED from DXPCTSUG-FEDER unha maneira de facer Europa (2012/273). The research leading to these results has received funding from the European Union's Seventh Framework Programme FP7/REGPOT-2012-2013.1 under grant agreement n° 316265, BIOCAPS. This document reflects only the author's views and the European Union is not liable for any use that may be made of the information contained herein. The authors also acknowledge PhD Grant of Ana Margarida Sousa SFRH/BD/72551/2010.

References

1. Lopez, D., Vlamakis, H., Kolter, R.: Biofilms. *Cold Spring Harb. Perspect Biol.* 2(7), a000398 (2010)
2. Yang, L., et al.: Current understanding of multi-species biofilms. *Int. J. Oral Sci.* 3(2), 74–81 (2011)
3. Stewart, P.S., Franklin, M.J.: Physiological heterogeneity in biofilms. *Nature Reviews Microbiology* 6(3), 199–210 (2008)
4. Lourenço, A., et al.: Minimum information about a biofilm experiment (MIABiE): standards for reporting experiments and data on sessile microbial communities living at interfaces. *Pathogens and Disease* (in press, 2014)
5. Lourenço, A., et al.: BioOmics: a Web platform for the systematic and standardized collection of high-throughput biofilm data. *PLoS One* 7(6), e39960 (2012)
6. Smith, B., et al.: The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* 25(11), 1251–1255 (2007)
7. Stevens, R., Goble, C.A., Bechhofer, S.: Ontology-based knowledge representation for bioinformatics. *Brief Bioinform.* 1(4), 398–414 (2000)
8. Buckingham-Meyer, K., Goeres, D.M., Hamilton, M.A.: Comparative evaluation of biofilm disinfectant efficacy tests. *Microbiol. Methods* 70(2), 236–244 (2007)
9. Soldatova, L.N., King, R.D.: An ontology of scientific experiments. *J. R. Soc. Interface* 3(11), 795–803 (2006)