# A harmonised vocabulary for communicating and interchanging Biofilms experimental results

**Ana Margarida Sousa[1], Maria Olívia Pereira[1], Nuno F. Azevedo[2], Anália Lourenço[3,1*]**

[1]CEB- Centre of Biological Engineering, LIBRO - Laboratório de Investigação em Biofilmes Rosário Oliveira, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

[2]LEPABE Dep. of Chemical Engineering, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal

[3]ESEI - Escuela Superior de Ingeniería Informática, Edificio Politécnico, Campus Universitario As Lagoas s/n, Universidad de Vigo, 32004 Ourense, Spain

**Summary**

Biofilm studies are at the crossroads of Biology, Chemistry, Medicine, Material Science and Engineering, among other fields. Data harmonisation in Biofilms is therefore crucial to allow for researchers to collaborate, interchange, understand, and replicate studies at an inter-laboratory and inter-domain scale. The international Minimum Information About a Biofilms Experiment initiative has prepared a set of guidelines for documenting biofilms experiments and data, namely the minimum information checklist. This paper goes a step forward and describes a new ontology for the broad description of biofilm experiments and data. In such an interdisciplinary context we chose to rely on a common integration framework provided by a foundational ontology that facilitates the addition and extension of various sub-domain modules, and the consistent integration of terminology extracted from several existing ontologies, e.g. EXPO and ChEBI. The community is participating actively in the production of this resource, and it is already used by public biofilms-centred databases, such as BiofOmics, and bioinformatics tools, such as the Biofilms Experiment Workbench. This practical validation serves the purpose of disseminating the controlled vocabulary among researchers and identifying current limitations, glitches, and inconsistencies. Information branches will be added, extended or refactored according to user feedback and group discussions.

## 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].

*To whom correspondence should be addressed. Email: analia@uvigo.es

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 ecological niche of the community and the relationships established among the constituent species, microorganisms have different metabolic, physiologic and genetic profiles that lead to unique biofilm signatures [3].

The major challenge to be faced in documenting biofilm experiments and results is the complexity and variability of biofilm studies (Figure 1). Researchers may study the structure and metabolism of biofilms, address the identification of the genes or gene products that contribute to particular biofilm formation, the role of the biofilms in the development of antimicrobial resistance, the communication systems orchestrating the dynamic of the community, or the ecological relationship prevailing within the biofilm, just to name a few subjects of interest.

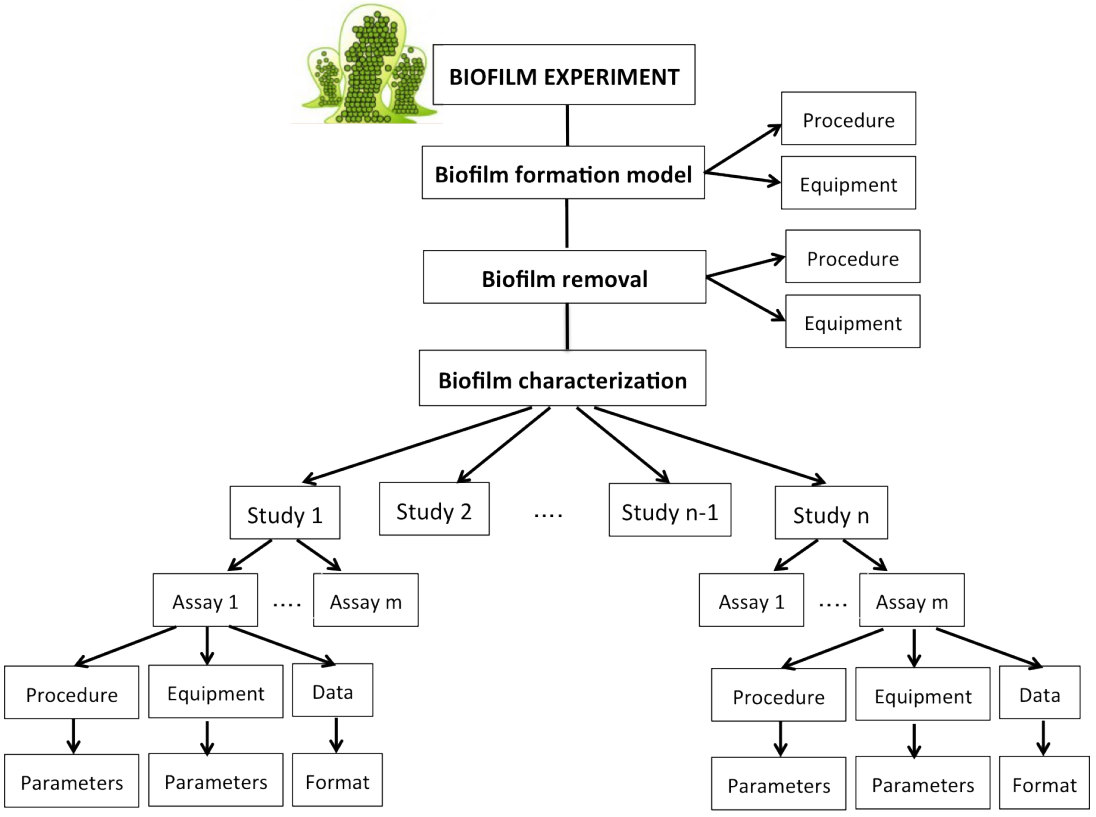


**Figure 1: A general and abstract view of a biofilm study. A biofilm study consists of a number of different disciplinary studies, e.g. microscopy, proteomics, metabolomics, and transcriptomics analyses. Those studies can involve numerous assays and each assay is composed by an experimental procedure, with a set of parameters, equipment, which also have parameters, and generates data in particular format according the kind of assay performed.**

Consequently, biofilm studies often encompass interdisciplinary approaches to the characterisation of the structure and activity of the community, such as microscopy, flow cytometry, antimicrobial susceptibility, metabolomics, proteomics and transcriptomics techniques. This concerted effort of study generates outputs that vary greatly in nature, structure, data volume and interpretation (Table 1). The management of such heterogeneous metadata and data is

complicated and raises several quality issues, such as lack of data reproducibility, scarcity of standardised protocols, poor data quality and incomplete data sets, which affects significantly the quality of the biofilms results being published [4, 5].

**Table 1: Typical types of biofilm data and techniques.**

| 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 |

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.

To cope with quality issues, the international Minimum Information About a Biofilms Experiment (MIABiE) initiative (http://miabie.org) is working on the harmonisation of biofilms procedures and associated data. In particular, this initiative has proposed guidelines about the minimum information that needs to be reported to guarantee the interpretability and independent verification of experimental results, and their integration with knowledge coming from other fields [6]. The creation of controlled vocabulary in support of the description of biofilms studies falls into the rationale of this harmonisation effort and seeks to support inter-laboratory data sharing, functional enrichment, and data mining.

The aims of this paper are to introduce the new MIABiE-compliant biofilms ontology, to describe on-going developments and validation, and to explain how the community is participating in the project. The basic idea is to develop a practical and semantically structured vocabulary for biofilm studies. Community resources and tools, such as the BiofOmics database [7] and the Biofilms Experiment Workbench [8], are endorsing this vocabulary and thus, users are becoming familiar with the resource in a seamless and practical way. As a side effect, the vocabulary is put into test under various scenarios of application so that different sub-communities may discuss the requirements of particular research interests and bring forward their recommendations. Likewise, the adequateness and extensibility of this ontology is being discussed within MIABiE workgroups.

The next sections are organised as follows: *Methods* section presents the main design principles of the ontology; *Results & Discussion* section presents the structure and contents of the ontology, some examples of the annotation of biofilm studies and collaborative efforts with other ontologies; *Ongoing Work* section describes work under development; and Conclusions section has final remarks about the work and community feedback.

## 2 Methods

There is no standardised methodology for building controlled vocabularies. However, the Open Biological and Biomedical Ontologies (OBO) foundry has introduced some useful guidelines and principles regarding the different stages of the ontology development life-cycle [9]. In particular, the design of the Biofilms ontology was based on the criteria below:

- the vocabulary is restricted to the biofilm knowledge domain and, therefore, it contains just model concepts and relations that are specific to the representation of biofilm data;
- the vocabulary should encompass all the information previously identified by MIABiE as essential to the comprehensive report of experimental findings [6];
- the vocabulary should be used for annotating data in databases and tools and for textual documentation, i.e. it should be understandable to people and unambiguously interpreted by software;
- the vocabulary should cope with new experimental studies, i.e. as new devices, techniques or applications arise, it should be possible to integrate new information branches without affecting the ontology structure;
- any biofilm experiment should be comprehensively described by a combination of terms;
- whenever possible, terms should be cross-referenced to entries on other controlled vocabularies dedicated to the domains contributing to biofilms studies.

Then, the ontology production life-cycle described by Steven and co-workers [10] was implemented as follows: (1) definition of the domain and scope of the vocabulary; (2) knowledge acquisition based on literature review and meetings with domain experts; (3) re-use of related and harmonised vocabulary; (4) knowledge representation using the OBO tool; and, (5) assessment of the correctness, accuracy and usefulness of the vocabulary.

Top-down and bottom-up approaches were combined in the construction of the ontology. First, the top-down approach led to the insertion of the Biofilms ontology 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, regardless the research field. As such, the Biofilms ontology may focus on the specificities of biofilms studies, and delegate 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 suitable relations among them. Since Biofilms are a multidisciplinary knowledge field, many concepts are cross-linked to ontologies in related domains, such as the Gene Ontology (GO) [11], the Functional Genomics Ontology (FuGO) [12], the Microarray Gene Expression Data Ontology (MO) [13], the Protein ontology (PRO) [14] and the Chemical Entities of Biological Interest (ChEBI) [15], avoiding term duplication and enforcing data interoperation across platforms and resources.

# 3 Results and discussion

## 3.1 Ontology structure

The Biofilms ontology is represented as a direct acyclic graph with four top level branches: 'study design', 'information about biofilm recovery', 'information about biofilm formation', and 'information about biofilm characterisation' (Figure 2).

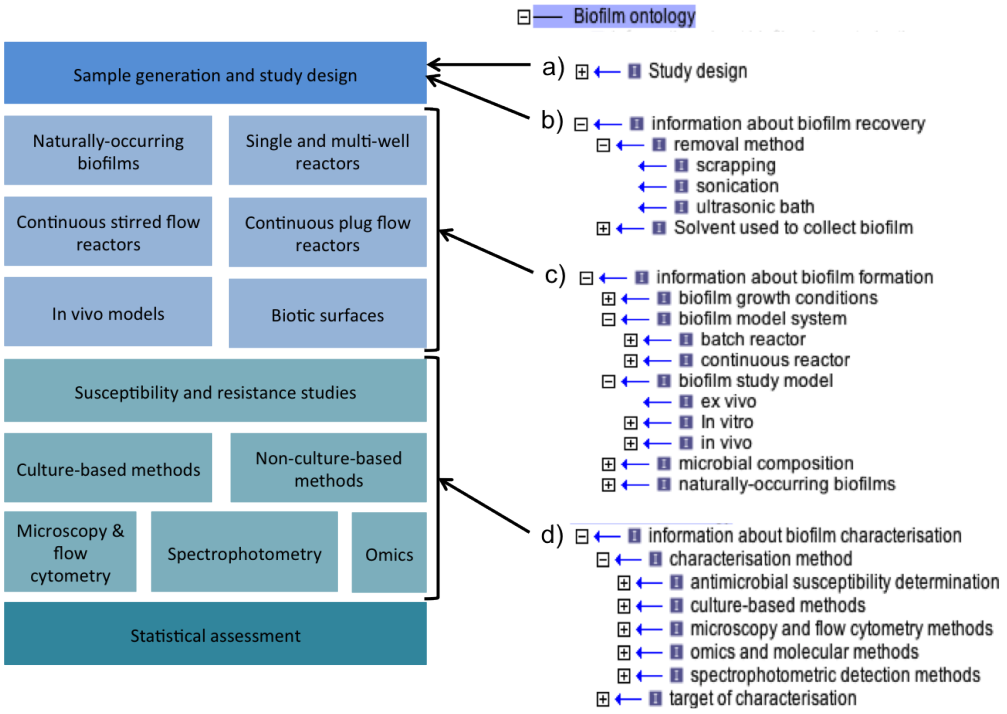


**Figure 2: The structure of the Biofilms ontology. The top-level branches (right side of the figure) are in compliance with the basic modules of information proposed by the MIABiE initiative (left side of the figure).**

The declaration of concepts consists in a unique identifier, a common name, the formal definition of the term and relevant references, alternative (exact or related) names, relation to other concepts, and links to other semantic sources, if applicable. By using a direct acyclic graph to structure the terms and relations, it is allowed that a term may be "child" of more than one "parent" term. For instance, the terms 'continuous', 'discontinuous', 'stepwise' and 'undefined' are "children" terms of the terms 'mode of stress exposure', 'temperature' and 'pH'.

## 3.2 Ontology contents

Ontology development started by defining the fundamental concepts about experiment design, the 'study design' branch, and the key terms related to three of the main stages of a biofilm experiment, most notably the 'information about biofilm formation', the 'information about biofilm recovery' and the 'information about biofilm characterisation' branches.

### 3.2.1 Study design

This branch includes terms related to the main characteristics of the study, including person(s) of contact, author affiliations, objective(s) of the study, type of experiment (one-factor, two-factor or multi-factor experiment), date of execution of the experiments, associated publication (if any). The design of the study is essential for researchers to understand the type of biofilm study at hands and, in particular, the scientific methodologies used to achieve the objective(s) stipulated. This kind of information is transversal to scientific studies in general and as so, the terms of this branch were cross-linked to upper ontologies of general knowledge (see section 3.3).

### 3.2.2 Information about biofilm formation

Following the typical biofilm protocol (Figure 1), the first level of information concerns the description of biofilm formation addressed by the terms in the branch 'information about biofilm formation'. This branch was decomposed into 'biofilm study model', 'biofilm model system', 'naturally-occurring biofilm', 'biofilm growth conditions', and 'microbial composition'.

The branch 'biofilm study model' describes the experimental setup used to form biofilms, such as *in vitro*, *ex vivo* and *in vivo* techniques [16]. There are numerous devices, apparatus and system configurations that can be used and that have tremendous impact on the biofilm structure and ecology and hence, are reflected on the results obtained. The 'biofilm model system' branch encompasses all experimental systems that can be used to form biofilms *in vitro* and *ex vivo*, including devices, apparatus and system configurations properly separated by their operation mode, continuous and batch reactors [17]. For example, microtiter plate-based model systems (such as the 96-, 48-, 24-, 12-, and 6-well plates, and the Calgary Biofilm Device), flow cell reactors, rotating disk reactors, propeller reactors, and drip-flow plate reactors.

The 'biofilm growth conditions' branch gathers together the terms commonly used in the characterisation of the conditions supporting biofilm development. Conditions such as 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. The description of these conditions is extremely important since any minor alteration can have a great impact on biofilm features and bias the results obtained.

Studies may also aim to study biofilms in natural habitats. For instance, biofilms formed in tubes of industrial processes, drinking-water distribution systems, indwelling devices, and teeth or human tissues. These biofilms cannot be characterised by general and well-defined human-controlled conditions and so the set of terms used in their particular description has been aggregated into the 'naturally-occurring biofilms' branch.

Last, the 'microbial composition' branch concerns the number of species and strains present in the biofilms. For instance, a biofilm composed just by one, two or more species is named of mono-, dual or multi-species, respectively. This kind of information is important for the characterisation of biofilm ecology, namely to species and strains distribution, establishment

of relations of competition, antagonism, symbiosis among organisms, and definition of distinct physiological niches [2, 3].

#### 3.2.3 Information about biofilm recovery

Typically, biofilm characterisation methods and techniques, with the exception of *in situ* techniques, require the removal of the biofilms from the surfaces where they were formed and the separation of cells from the surrounding exopolysaccharide matrix. The quality of the biofilm sampling in biofilm detachment and cell separation is of upmost importance because it may alter the physiological state of cells and bias the results of the experiment. So, the description of such sampling should be complete and comprehensible. This branch of the ontology describes the existing methods and techniques, such as sonication, scrapping and ultrasonic bath. It also describes common operation conditions, such as the device used to remove the biofilm (for example, the power of the sonicator and the size of the scrappers), period of time of biofilm removal (for example, number and duration of sonication cycles) and solvent used to collect the biofilm cells (for example, water, phosphate buffer or other).

#### 3.2.4 Information about biofilm characterisation

This branch of the ontology describes the methods used to characterise the biofilm, namely: culture-based methods, such as the counting of viable cells and colony morphology characterisation; non culture-based methods, such as the DAPI, CV and FISH analytical methods; microscopy (e.g. SEM, TEM and confocal techniques) and flow cytometry methods; spectrophotometric methods, such as XTT, ATP detection and Lowry protein assay; ”omics” and molecular methods, such as transcriptomics, metabolomics, proteomics, genomics; and antimicrobial susceptibility testing, such as the determination of the minimum inhibitory concentration (MIC), the minimum bactericidal concentration (MBC) and the minimum biofilm eradication concentration (MBEC). The definition of techniques developed by other research areas is retrieved from domain-specific ontologies and controlled vocabularies. For example, the description of the mass spectrometry analysis of biofilm-derived samples should encompass spectrum generation and spectrum interpretation, following the directives of the Proteomic Standards Initiative [18].

### 3.3 Extending and harmonising controlled vocabularies and ontologies

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. By integrating some of those ontologies, the Biofilms ontology may focus on the specificities of biofilms studies and delegate general experimental characterisation (Figure 3).

None of the upper ontologies is an ideal representation of general knowledge and, therefore, there must be a compromise between the degree of ”imperfection” of the upper ontology and the needs of our ontology as practical domain ontology. Here, the Suggested Upper Merged

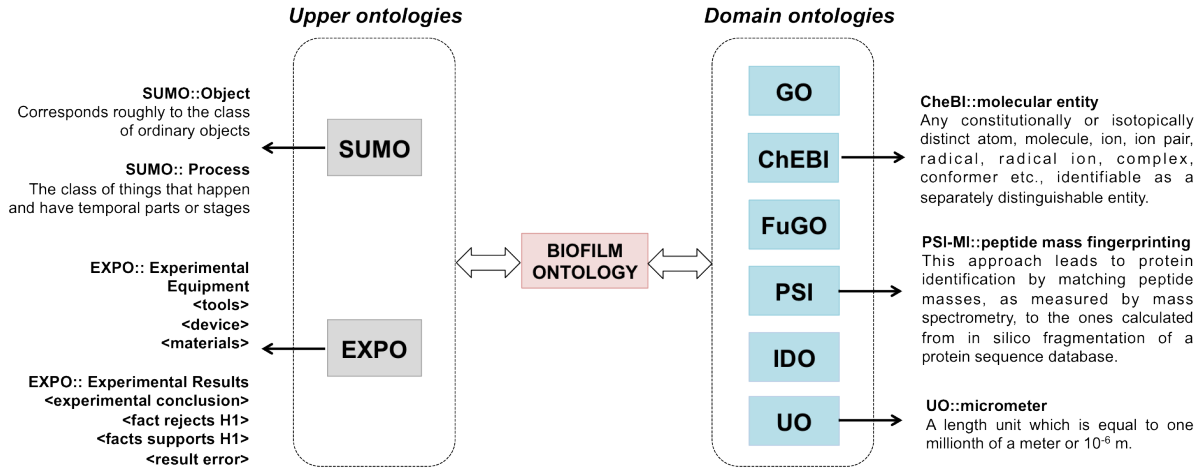


**Figure 3: The interoperation of the Biofilms ontology with domain ontologies (at the right) and upper ontologies (at the left) and some examples of terms reused. GO – Gene Ontology; ChEBI – Chemical Entities of Biological Interest; FuGO – Functional Genomics Ontology; PSI – Protein Standards Initiative; IDO – Infection and Disease Ontology; UO – Units of measurement Ontology; SUMO – Suggested Upper Merged Ontology; EXPO – Experiment Ontology.**

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 and engineering).

The Biofilms ontology 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 [19]. For example, any experiment has a goal ('EXPO:ExperimentalGoal') that can be confirmed ('EXPO:ConfirmGoal'), explained ('EXPO:ExplainGoal'), investigated ('EXPO:InvestigateGoal') or computed ('EXPO:ComputeGoal'). Any experiment aims to test a hypothesis ('EXPO:ExperimentalHypothesis'), meaning that there is a research hypothesis ('EXPO:ResearchHypothesis') that results may confirm ('EXPO:FactSupport') or reject ('EXPO:RejectSupport').

The integration and cooperation with other controlled vocabularies and ontologies is not limited to generic or top-level ontologies. Since Biofilms are a multidisciplinary knowledge field, many concepts are cross-linked to ontologies in related domains, avoiding overlapping and enforcing data interoperation across platforms and resources, in benefit of the broad research community. Therefore, data coming from other knowledge domains, such as flow cytometry, proteomic techniques or microarrays, are annotated according to the data standards of the respective consortia. A summary of the external resources that are considered to be the most important to Biofilms is shown in Table 2.

doi:10.2390/biecoll-jib-2014-249

**Table 2: Typical types of biofilm data and techniques.**

| Ontology / controlled vocabulary | Source | Reference |
|---|---|---|
| Chemical entities of biological Interest (CHEBI) | http://obo.cvs.sourceforge.net/obo/obo/ontology/chemical/chebi.obo | [15] |
| Infectious Disease Ontology (IDO) | http://www.bioontology.org/wiki/index.php/Infectious_Disease_Ontology | [20] |
| Colony morphology characterization (CMO) | http://mibbi.sourceforge.net/projects/MIABiE.shtml | [21] |
| Gene ontology (GO) | http://obo.cvs.sourceforge.net/obo/obo/ontology/genomic-proteomic/gene_ontology.obo | [11] |
| Functional Genomics Investigation Ontology (FuGO) | http://sourceforge.net/projects/fugo/ | [12] |
| MALDI imaging ontology (IMS) | http://www.maldi-msi.org/download/imzml/imagingMS.obo | [22] |
| PSI-Molecular Interactions (MI) | http://obo.cvs.sourceforge.net/obo/obo/ontology/genomic-proteomic/protein/psi-mi.obo | [23] |
| PSI-Protein modifications (MOD) | http://psidev.cvs.sourceforge.net/viewvc/psidev/psi/mod/data/PSI-MOD.obo | [24] |
| PSI-Mass Spectrometry (MS) | http://psidev.cvs.sourceforge.net/viewvc/psidev/psi/psi-ms/mzML/controlledVocabulary/psi-ms.obo | [18] |
| PSI-Sample Processing and Separations (SEP) | https://psidev.svn.sourceforge.net/svnroot/psidev/psi/sepcv/trunk/sep.obo | [22] |
| PRIDE controlled vocabulary | http://code.google.com/p/ebi-pride/source/browse/trunk/pride-core/schema/pride_cv.obo | [25] |
| Protein ontology (PRO) | http://obo.cvs.sourceforge.net/obo/obo/ontology/genomic-proteomic/pro.obo | [14] |
| Phenotypic and Trait Ontology (PATO) | http://obo.cvs.sourceforge.net/obo/obo/ontology/phenotype/unit.obo | [26] |

### 3.4 Ontology validation

The primary goal of a controlled vocabulary or ontology is to be a means to create coherent, machine-readable annotations in support of resource catalogues, information standards, collaborative infrastructures, and software.

To ensure good coverage of the vocabulary involved in the description of common biofilm study stages, factors and variables, several studies of distinct research fields were semantically annotated with the Biofilms ontology (Figure 4). By making the exercise of annotating real-world biofilm studies, we could assess the correctness and the usability of the terms included in the ontology from the point of view of both resource annotators (i.e. the vocabulary at their disposal to document the experiment) and users (i.e. the extent of details that could be expected). Next, the ontology was presented and discussed with field experts at international conferences.

Tests and discussion with other researchers evidenced the inconsistent organisation and description of the experimental protocols related to biofilm formation, recovery and characterisation in the literature. Biofilm protocols include a very broad range of activities and their description is often ambiguous. In fact, numerous details of the experimental procedures are, in general, not available in scientific publications (namely in "Materials and Methods" section, as they should), being only known by the author of the study and his research group/lab. There is a loss of relevant data to the reproduction and comparison of experiments. So, the most efficient "recipe" to harmonise Biofilms vocabulary and produce comprehensive documentation is to combine the use of Biofilms ontology with the compliance with MIABiE checklists, i.e. systematic and consolidated level of detail. In particular, the Biofilms ontology can make a difference in the computerised management and analysis of experiments and experimental results, previous and complimentary to scientific publication.

Therefore, the next level of validation addressed the use of the Biofilms ontology in publicly available and community-driven resources and tools. Specifically, the BiofOmics database and the Biofilms Experiment Workbench have endorsed the ontology and are orchestrating its revision (Figures 5 and 6), providing us feedback on the experience of both data submitters and database curators. In particular, Biofilms Experiment Workbench enables the editing of the vocabulary by the user so that the specifics of a sub-domain of study may be covered, and the BiofOmics curators are responsible for curating the candidate terms and creating an up-to-date version of the ontology.

## 4 Ongoing work

We reckon that the development of the Biofilms ontology is at its beginning. The practical approach presented here should continue in order to extend the definition of the procedures, methods and techniques to areas not yet covered. So, this work is under regular discussion with MIABiE members and other domain experts. For example, the creation of modular ontologies in the branch 'information about biofilm characterisation' is currently under debate. In particular, the integration of terminology from other domains is not straightforward. Methods such

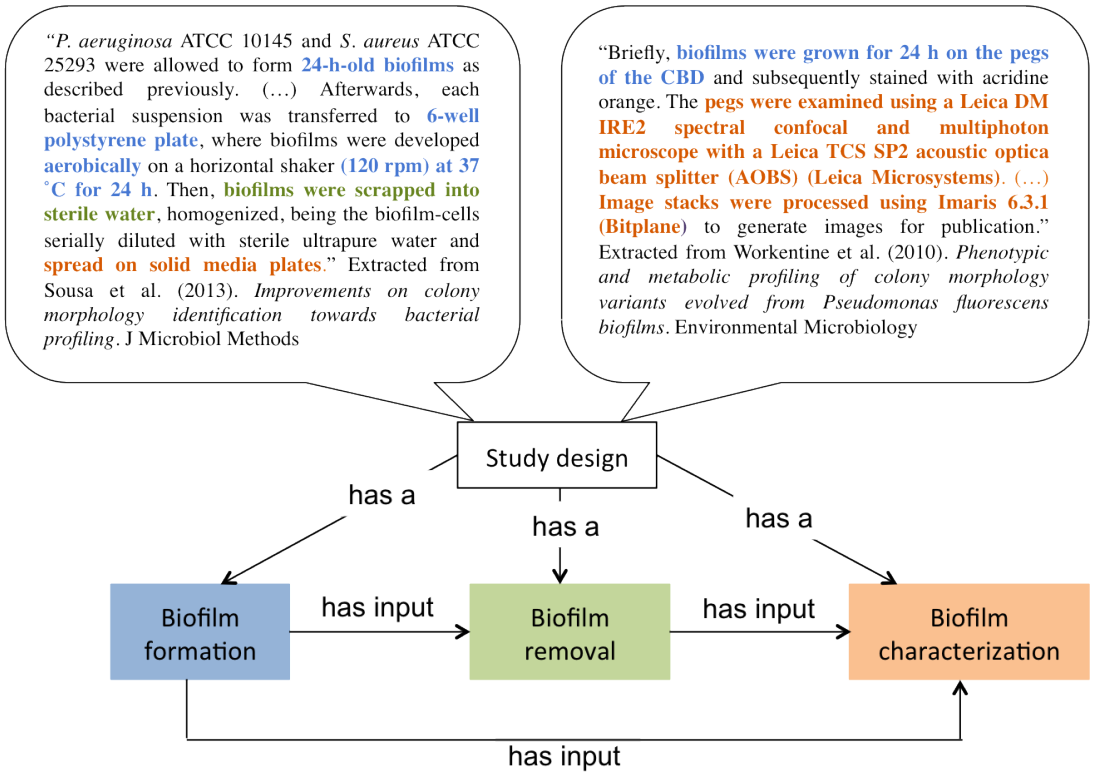


**Figure 4: Information model for semantic annotation of biofilm studies using the new Biofilms ontology.**

as antimicrobial susceptibility testing are crucial to Biofilms research, in particular for clinical and food related studies, but they are not Biofilms specific. The problem is that, as far as we know, no ontology or controlled vocabulary is tackling antimicrobial susceptibility testing at the moment. Therefore, the MIABiE is considering the option of engaging in the development of such a non-Biofilms exclusive ontology, as a module of the Biofilms ontology, to meet the requirements of the domain.

Participation in the development of the Biofilms ontology is not restricted to MIABiE work-groups though. We feel that the Biofilms community should participate actively in the validation and maintenance of the ontology. Researchers are invited to propose the addition, amendment and even removal of concepts. Feedback from collaborating databases or other semantic resources are also welcome.

The Biofilms ontology is being iteratively revised, yielding at least one version per year, being in the current version 1.2. End-user query and annotation tools are being developed to bring the process closer to Biofilms researchers and enable collaborative annotation.

## 5 Conclusions

The Biofilms community is investing in data harmonisation, namely minimum information standards, file formats and controlled vocabulary. Here, the first stage of development of the

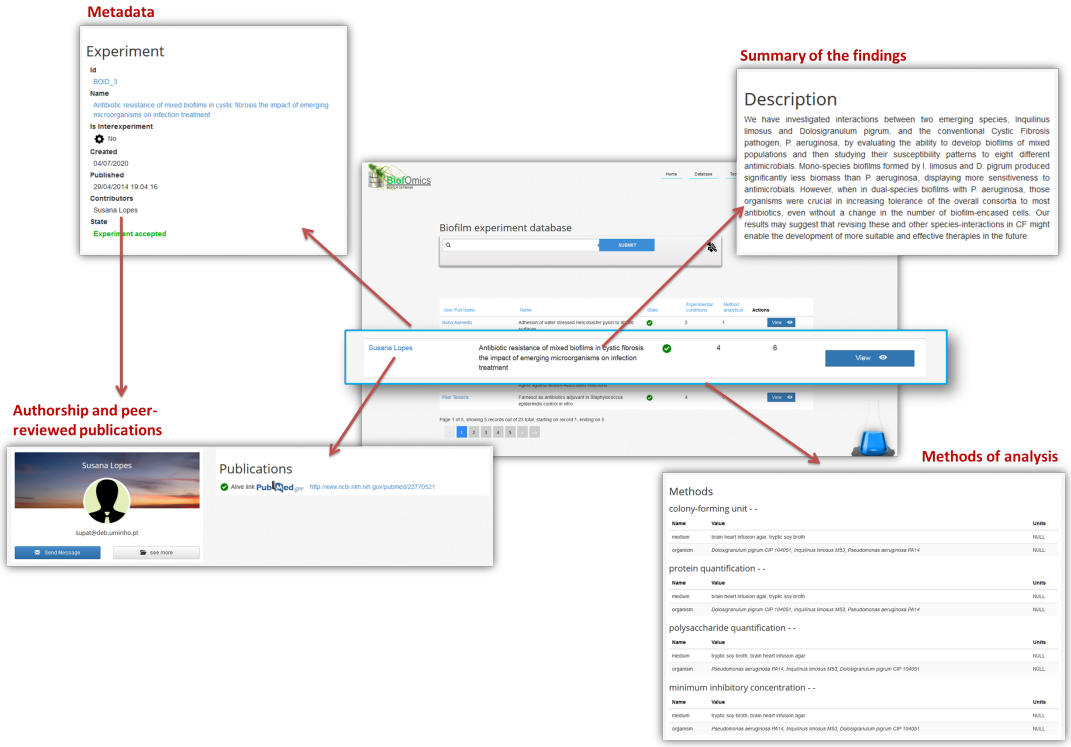


**Figure 5: Example of the annotation of BiofOmics database records with the Biofilms ontology. The annotation includes the study design, authorship information and any related peer reviewed publications, and metadata and data collection.**

Biofilms ontology was presented. Proper semantic annotation of biofilm studies is crucial to link in an unambiguous way the unstructured descriptions found in scientific articles and thus, to create a global understanding of the results that are being obtained. Notably, the community benefits greatly from the existence of public databases on experimental results and bioinformatics tools, supporting experimental data processing and analysis.

The new Biofilms ontology covers all the steps of a typical biofilm experiment workflow, including 'information about biofilm formation', 'information about biofilm recovery' and 'information about biofilm characterisation'. The added value of this ontology lies in focusing on Biofilms-specific vocabulary and delegating non-specific vocabulary to the ontologies of the corresponding domain. Also, the ontology 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 actively collaborating in the population and update of the ontology.

## Acknowledgements

The financial support from IBB-CEB, Fundao para a Ciłncia e Tecnologia (FCT), European Community fund FEDER, trough Program COMPETE, in the ambit of the projects PEst-OE/EQB/LA0023/2013 and PTDC/SAUSAP/113196/2009/FCOMP-01-0124-FEDER-016012

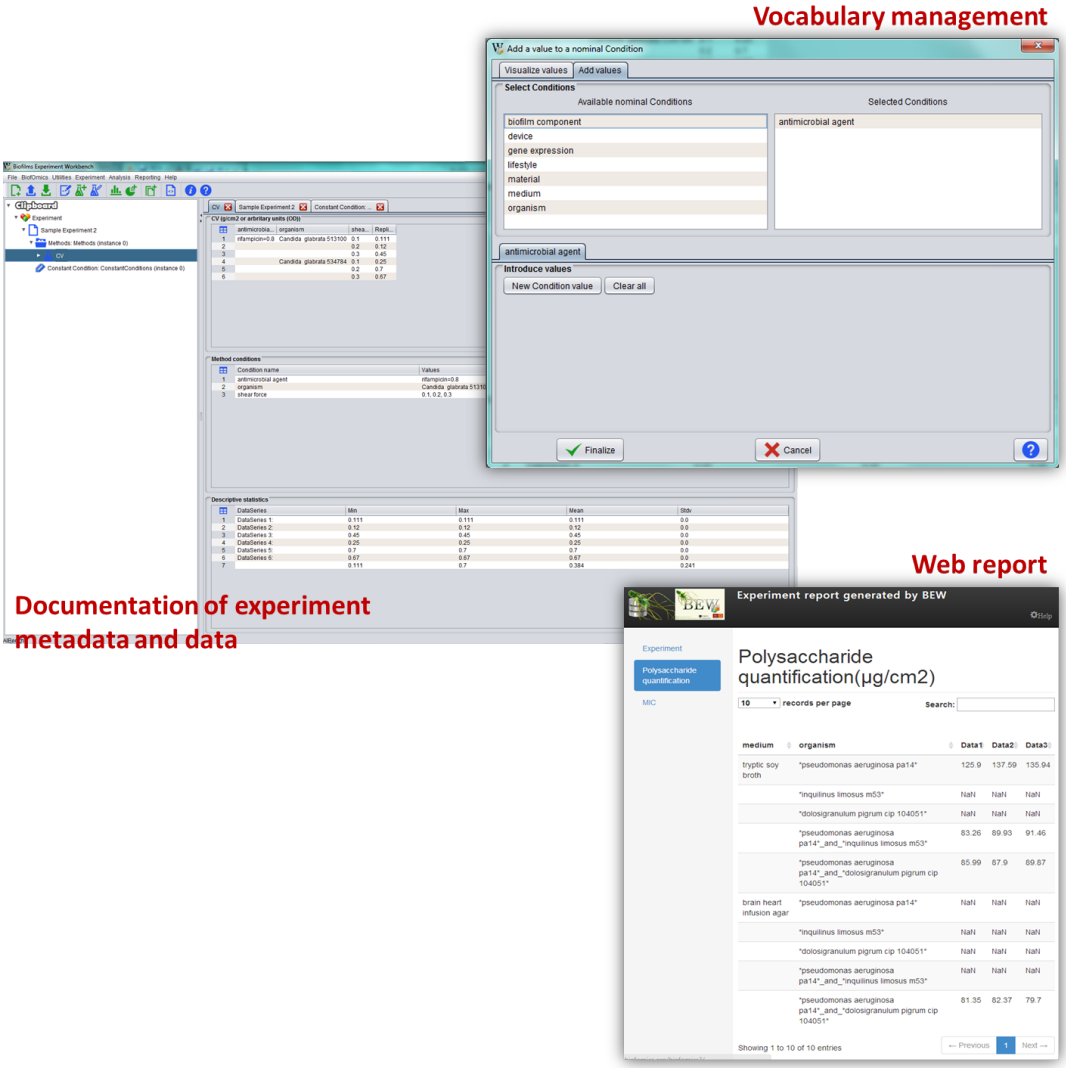


**Figure 6: Example of an experiment report generated by the Biofilms Experiment Workbench and supported by the Biofilms ontology.**

and Ana Margarida Sousa PhD Grant (SFRH/BD/72551/2010), European Union Seventh Framework Programme [FP7/REGPOT-2012-2013.1] [grant number 316265], BIOCAPS and the Agrupamento INBIOMED from DXPCTSUG-FEDER unha maneira de facer Europa (2012/273) are gratefully acknowledged.

## References

[1] D. Lopez, H. Vlamakis, R. Kolter. Biofilms. *Cold Spring Harb Perspect Biol*, 2:a000398, 2010.

[2] L. Yang, Y. Liu, H. Wu, N. Hoiby, S. Molin, Z. Song. Current understanding of multi-species biofilms. *International Journal of Oral Science*, 3:74–81, 2011.

[3] P.S. Stewart, M.J. Franklin. Physiological heterogeneity in biofilms. *Nature Reviews Microbiology* 6:199–210, 2008.

[4] A.M. Sousa, A. Ferreira, N.F. Azevedo, M.O. Pereira, A. Lourenco. Computational approaches to standard-compliant biofilm data for reliable analysis and integration. *Journal of Integrative Bioinformatics*, 9:203, 2012.

[5] Y. Huang, R. Gottardo. Comparability and reproducibility of biomedical data. *Briefings in Bioinformatics*, 14:391–401, 2013.

[6] A. Lourenço, T. Coenye, D.M. Goeres, G. Donelli, A. Azevedo, H. Ceri, F.L. Coelho, H. Flemming, T. Juhna, S.P. Lopes, R. Oliveira, A. Oliver, M.E. Shirtliff, A.M. Sousa, P. Stoodley, M.O. Pereira, N.F. Azevedo. Minimum information about a biofilm experiment (MIABiE): standards for reporting experiments and data on sessile microbial communities living at interfaces. *Pathogens and Disease*, 70:250–256, 2014.

[7] A. Lourenço, A. Ferreira, N. Veiga, I. Machado, M.O. Pereira, N.F. Azevedo. BiofOmics: a Web platform for the systematic and standardized collection of high-throughput biofilm data. *PloS one*, 7:e39960, 2012.

[8] G. Rodríguez, D. Glez-Peña, N.F. Azevedo, M.O. Pereira, F. Fdez-Riverola, A. Lourenço. BEW: Bioinformatics Workbench for Analysis of Biofilms Experimental Data. In *8th International Conference on Practical Applications of Computational Biology and Bioinformatics (PACBB 2014)*, Saez-Rodriguez, J., Rocha, M. P., Fdez-Riverola, F., De Paz Santana, J. F., Eds., Springer International Publishing, 294:49–56, 2014.

[9] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L.J. Goldberg, K. Eilbeck, A. Ireland, C.J. Mungall, the OBI Consortium, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S.A. Sansone, R.H. Scheuermann, N. Shah, P. Whetzel, S. Lewis. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25:1251–1255, 2007.

[10] R. Stevens, C.A. Goble, S. Bechhofer. Ontology-based knowledge representation for bioinformatics. *Briefings in Bioinformatics*, 1:398-414, 2000.

[11] Gene Ontology Consortium. Creating the gene ontology resource: design and implementation. Genome Research, 11:1425–1433, 2001.

[12] P.L. Whetzel, R.R. Brinkman, H.C. Causton, L. Fan, D. Field, J. Fostel, G. Fragoso, T. Gray, M. Heiskanen, T. Hernandez-Boussard, N. Morrison, H. Parkinson, P. Rocca-Serra, S.A. Sansone, D. Schober, B. Smith, R. Stevens, C.J. Stoeckert Jr., C. Taylor, J. White, A. Wood, FuGO Working Group. Development of FuGO: an ontology for functional genomics investigations. *Omics: A Journal of Integrative Biology*, 10:199–204, 2006.

[13] P.L. Whetzel, H. Parkinson, H.C. Causton, L. Fan, J. Fostel, G. Fragoso, L. Game, M. Heiskanen, N. Morrison, P. Rocca-Serra, S.A. Sansone, C. Taylor, J. White, C. J. Stoeckert Jr. The MGED Ontology: a resource for semantics-based description of microarray experiments. *Bioinformatics*, 22:866–873, 2006.

[14] D.A. Natale, C.N. Arighi, J.A. Blake, C.L. Bult, K.R. Christie, J. Cowart, P. D'Eustachio, A.D. Diehl, H.J. Drabkin, O. Helfer, H. Huang, A.M. Masci, J. Ren, N.V. Roberts, K. Ross, A. Ruttenberg, V. Shamovsky, B. Smith, M. Shruti Yerramalla, J. Zhang, A. Al-Janahi, I. Celen, C. Gan, M. Lv, E. Schuster-Lezell, C.H. Wu. Protein Ontology: a controlled structured network of protein entities. *Nucleic Acids Research*, 42:D415–421, 2014.

[15] K. Degtyarenko, P. Matos, M. Ennis, J. Hastings, M. Zbinden, A. McNaught, R. Alcantara, M. Darsow, M. Guedj, M. Ashburner. ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Research*, 36:D344–350, 2008.

[16] T. Coenye, H.J. Nelis. *In vitro* and *in vivo* model systems to study microbial biofilm formation. *Journal of Microbiological Methods*, 83:89–105, 2010.

[17] K. Buckingham-Meyer, D.M. Goeres, M.A. Hamilton. Comparative evaluation of biofilm disinfectant efficacy tests. *Journal of Microbiological Methods*, 70:236–244, 2007.

[18] G. Mayer, L. Montecchi-Palazzi, D. Ovelleiro, A.R. Jones, P.-A. Binz, E.W. Deutsch, M. Chambers, M. Kallhardt, F. Levander, J. Shofstahl, S. Orchard, J.A. Vizcaino, H. Hermjakob, C. Stephan, H.E. Meyer, M. Eisenacher. The HUPO proteomics standards initiative- mass spectrometry controlled vocabulary. *Database: The Journal of Biological Databases and Curation*, 2013:bat009, 2013.

[19] L.N. Soldatova, R.D. King. An ontology of scientific experiments. *Journal of the Royal Society*, Interface / the Royal Society, 3:795–803, 2006.

[20] A. Goldfain, B. Smith, L.G. Cowell. Towards an ontological representation of resistance: the case of MRSA. *Journal of Biomedical Informatics*, 44:35–41, 2011.

[21] A.M. Sousa, A. Lourenço, M.O. Pereira. MorphoCol: a powerful tool for the clinical profiling of pathogenic bacteria. In *Advances in Intelligent and Soft Computing*, Springer, 154:181–187, 2012.

[22] G. Mayer, A.R. Jones, P.-A. Binz, E.W. Deutsch, S. Orchard, L. Montecchi-Palazzi, J.A. Vizcaino, H. Hermjakob, D. Oveillero, R. Julian, C. Stephan, H.E. Meyer, M. Eisenacher. Controlled vocabularies and ontologies in proteomics: Overview, principles and practice. *Biochimica et Biophysica Acta*, 1844:98–107, 2014.

[23] H. Hermjakob, L. Montecchi-Palazzi, G. Bader, J. Wojcik, L. Salwinski, A. Ceol, S. Moore, S. Orchard, U. Sarkans, C.n von Mering, B. Roechert, S. Poux, E. Jung, H. Mersch, P. Kersey, M. Lappe, Y. Li, R. Zeng, D. Rana, M. Nikolski, H. Husi, C. Brun, K. Shanker, Seth G. N. Grant, C. Sander, P. Bork, W. Zhu,, A. Pandey, A. Brazma, B. Jacq, M. Vidal, David Sherman, Pierre Legrain, G. Cesareni, I. Xenarios, D. Eisenberg, B. Steipe, C. Hogue, R. Apweiler. The HUPO PSI's Molecular Interaction format - a community standard for the representation of protein interaction data. *Nature Biotechnology*, 22:177–183, 2004.

[24] L. Montecchi-Palazzi, R. Beavis, P.A. Binz, R. J. Chalkley, J. Cottrell, D. Creasy, J. Shofstahl, S.L. Seymour, John Garavelli. The PSI-MOD community standard for representation of protein modification data. *Nature Biotechnology*, 26:864–866, 2008.

[25] L. Martens, H. Hermjakob, P. Jones, M. Adamski, C. Taylor, D. States, K. Gevaert, J. Vandekerckhove, R. Apweiler. PRIDE: the proteomics identifications database. *Proteomics*, 5:3537–3545, 2005.

[26] G.V. Gkoutos, E.C. Green, A.M. Mallon, A. Blake, S. Greenaway, J.M. Hancock, D. Davidson. Ontologies for the description of mouse phenotypes. *Comparative and Functional Genomics*, 5:545–551, 2004.