9th International Conference on Predictive Modelling in Food

Towards Community Driven Food Safety Model Repositories

Matthias Filter *, Carolina Plaza-Rodríguez, Christian Thoens, Annemarie Kaesbohrer
and Bernd Appel

Federal Institute for Risk Assessment, Berlin, Germany

Abstract

Transferring predictive microbial models from research into real world food manufacturing or risk assessment applications is still a challenge for members of the food safety modelling community. Such knowledge transfer could be facilitated if publicly available food safety model repositories would exist.

This research therefore aimed at identification of missing resources hampering the establishment of community driven food safety model repositories. Existing solutions in related scientific disciplines like Systems Biology and Data Mining were analyzed.

On the basis of this analysis, some factors which would promote the establishment of community driven model repositories were identified – among them: a standardized information exchange format for models and rules for model annotation. As a consequence a proposal for a Predictive Modelling in Food Markup Language (PMF-ML) together with a prototypic implementation on the basis of the Systems Biology Markup Language (SBML) has been developed. In addition the adoption of MIRIAM guidelines for model annotation is proposed. In order to demonstrate the practicability of the proposed strategy, existing predictive models previously published in the scientific literature were re-implemented using an open source software tool called PMM-Lab. The models are made publicly available in the first community Food Safety Model Repository called openFSMR (https://sites.google.com/site/openfsmr/).

This work illustrates that a standardized information exchange format for predictive microbial models can be established by adoption of resources from Systems Biology. Harmonized description and annotation of predictive models will also contribute to increased transparency and quality of food safety models.

* Corresponding author. Tel.: +49-30-18412-2209; fax: +49-30-18412-2952.
E-mail address: matthias.filter@bfr.bund.de

1. Introduction

Here we present the current status of an initiative aiming at the establishment of community driven and open Food Safety Model Repositories (FSMR). The motivation for this research arises from the understanding that there is a strong need for publicly available, collaboratively maintained and curated resources that allow software independent exchange of predictive models in the field of food safety modelling[1]. Currently there exist quite a number of different and sophisticated modelling and prediction tools (an online tool inventory is available at: http://sourceforge.net/p/microbialmodelingexchange/wiki/Tools/). However, there is no standardized way of exchanging information or data between these tools. For business operators and governmental agencies this lack of interoperability is a significant obstacle hampering the broad application of models generated by the scientific community. It also limits the applicability of existing software tools itself, averting efficient knowledge exchange and hindering quality control and transparent documentation of modelling results.

2. Results

Related efforts in the Systems Biology domain showcase that the successful establishment and long-term provisioning of model repositories depend on certain prerequisites:

1. Existence of a standardized file format for software-independent encoding of models
2. Existence of rules / guidance for the annotation of experimental data and / or models
3. Software tools supporting the adoption of file formats and annotation rules
4. Technical infrastructure for hosting web-based model repositories
5. Community support (creation, sharing and quality control of models)

This research therefore addressed each of these issues and makes newly developed resources publicly available in order to promote joint future developments.

2.1. Predictive Modelling in Food Markup Language (PMF-ML)

PMF-ML is a standard allowing the encoding of arbitrary food safety models into XML-formatted text files. PMF-ML adopts SBML, version 3[2]. It provides software developers with guidance on how to implement the SBML standard in the domain of food safety modelling. Such guidance is necessary in order to avoid e. g. misinterpretation of SBML concepts that have to be mapped to concepts used in the domain of food safety modelling. For example, a central SBML concept is "Species" which "... refers to a pool of entities that (a) are considered indistinguishable from each other for the purposes of the model, (b) may participate in reactions, and (c) are located in a specific compartment."[2]. PMF-ML clarifies that the SBML "Species" concept should be mapped to the "Organism" concept which is widely used in food safety models. In addition certain food safety domain-specific conventions are made inside the PMF-ML guidance document. For example, PMF-ML defines that certain model metadata, e.g. the reference to file(s) containing experimental data used for model generation, have to be provided inside the SBML annotation container using a newly introduced "pmf" namespace. The full specification is freely available at the website of the community project "OpenML for Predictive Modelling in Food": (http://sourceforge.net/projects/microbialmodelingexchange/).

2.2. Rules / guidance for annotation of models

The assignment of meaningful, standardized and exhaustive metadata for food safety models is one of the biggest challenges for food safety data scientists. In this domain there is currently no standard or guidance available defining how a predictive model should be annotated and which terms or description scheme should be used. We therefore

propose to adopt again an approach successfully applied in Systems Biology. Specifically we suggest adopting the "Minimal Information Required In the Annotation of Models" (MIRIAM) guidelines[3]. These guidelines demand that e.g. a model is clearly related to a single reference description, is instantiable in a simulation and that when it is instantiated, all results given in the reference description can be reproduced with the given model. Amongst others, MIRIAM requires to provide a model name, the name and contact information of the model creator(s), the date/time of model creation and a precise statement on the terms of use (license). Within the PMF-ML guidance document we provide recommendations on how to assign additional, non-compulsory metadata to food safety model files. For example we illustrate how the model quality, the range of validity as well as details on food matrices, organisms and experimental conditions should be provided.

2.3. Software tools supporting modellers in adopting standardized file formats and annotation rules

A critical factor influencing the rate of adoption of newly proposed standards and rules is the availability of software tools supporting them. By nature this is a chicken-and-egg dilemma specifically in the beginning. For this reason BfR funded the necessary technical developments for providing a reference implementation in the open source software PMM-Lab[4]. This software (PMM-Lab v1.07) has been extended such, that models generated with the PMM-Lab software can now be annotated according to the MIRIAM guidelines and stored as a PMF-ML file. In reverse PMM-Lab is capable of importing PMF-ML files which are compliant to the PMF guidelines. The PMM-Lab software can be downloaded freely at http://sourceforge.net/projects/pmmlab/.

2.4. Technical infrastructure for hosting web-based model repositories

As outlined in the introduction, the aim of this research has been the establishment of community-driven FSMRs. As a proof-of-concept implementation, we created a web-based model repository called openFSMR. This community driven information portal on food safety models is accessible at https://sites.google.com/site/openfsmr/. It has been designed with the primary objective to provide an inventory of food safety models available either as a PMF-ML formatted file or as part of publicly available software tools. Technically the openFSMR comprise of two components:

1. A tabular collection of models (openFSMR-DB) that includes relevant model metadata (stored as a Google Sheets tables).
2. A website displaying model metadata with sophisticated search and filtering functionalities (generated using Google Sites).

The openFSMR website also provides the opportunity to add new models into the openFSMR-DB (implemented with Google Forms). These new data entries will undergo a curation and quality control process which in the future should be performed by members of the "openFSMR editorial board" – see 2.5. The Google based infrastructure ensures that the openFSMR can be maintained and extended in a collaborative fashion, at minimal cost and in a highly transparent and flexible manner. As of September 2015 the majority of data inside openFSMR refer to models implemented into predictive microbial modelling tools.

2.5. Community support

The long term success of initiatives like this strongly depends on support generated by the food safety community itself. It is necessary to emphasize, that the desired increased acceptance of microbial modelling and risk assessment technologies by end users will only be achieved if software engineers, (risk) modellers, lab scientists, project managers, funding agencies and end users from food industry and governmental agencies jointly promote such community efforts. For example, it would be extremely beneficial if scientific journals would embrace the adoption of PMF-ML or MIRIAM guidelines for food safety models. Or, if newly initiated or ongoing software development projects also consider the technical support for the PMF-ML data exchange format. Even lab scientists could contribute by creating and sharing fully annotated experimental data sets and statistical models with the community.

3. Conclusion

This work illustrates that a standardized information exchange format for predictive microbial models can be established by adoption of resources from Systems Biology. Harmonized description and annotation of predictive models will contribute to increased transparency and quality of food safety models in the future. However, significant investments still have to be taken by the food safety modelling community to achieve their long term goals. For example, additional efforts are needed to critically review and improve the PMF-ML guidance document in itself (e.g. maintenance of controlled vocabularies, applicability to other model classes relevant for risk assessments, establishing support for SED-ML). There is also a strong need to adopt PMF-ML by existing software tools, so that PMF-ML import and export features will be available. Models from scientific publications have to be re-implemented, annotated and exported as PMF-ML files in order to make them publicly available through FSMRs and finally easily deployable by end users.

Acknowledgements

References

1.   Plaza-Rodriguez C, Thoens C, Falenski A, et al. A strategy to establish Food Safety Model Repositories. *Int J Food Microbiol*. 2015; 204: 81-90.
2.   Hucka M, Bergmann F, Hoops S, Keating SM, Sahle S and Wilkinson DJ. The Systems Biology Markup Language (SBML):
Language Specification for Level 3 Version 1 Core. 2009.
3.   Le Novere N, Finney A, Hucka M, et al. Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat Biotechnol*. 2005; 23: 1509-15.
4.   Filter M, Thöns C, Brandt J, et al. A community resource for integrated predictive microbial modelling (PMM-Lab). *5th International Workshop Cold Chain Management*. Bonn, Germany2013.