Harmonized terms, concepts and metadata for microbiological risk assessment models: the basis for knowledge integration and exchange

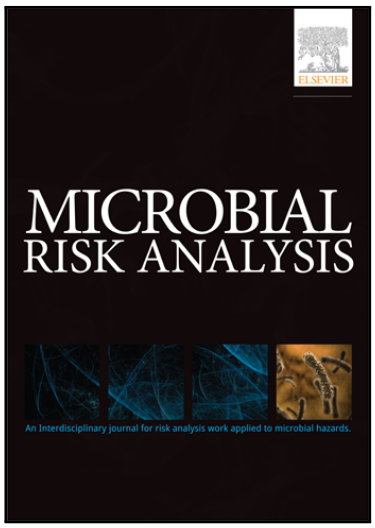
Ungaretti Haberbeck, Leticia; Rodríguez, Carolina Plaza; Desvignes, Virginie; Dalgaard, Paw; Sanaa, Moez; Guillier, Laurent; Nauta, Maarten; Filter, Matthias

Published in:
Microbial Risk Analysis

Link to article, DOI:
10.1016/j.mran.2018.06.001

Publication date:
2018

Document Version
Peer reviewed version



Harmonized terms, concepts and metadata for microbiological risk assessment models: the basis for knowledge integration and exchange

Leticia Ungaretti Haberbeck , Carolina Plaza Rodríguez , Virginie Desvignes , Paw Dalgaard , Moez Sanaa , Laurent Guillier , Maarten Nauta , Matthias Filter

PII: S2352-3522(18)30010-0
DOI: 10.1016/j.mran.2018.06.001
Reference: MRAN 56

To appear in: *Microbial Risk Analysis*

Received date: 23 February 2018
Revised date: 7 June 2018
Accepted date: 9 June 2018

Please cite this article as: Leticia Ungaretti Haberbeck , Carolina Plaza Rodríguez , Virginie Desvignes , Paw Dalgaard , Moez Sanaa , Laurent Guillier , Maarten Nauta , Matthias Filter , Harmonized terms, concepts and metadata for microbiological risk assessment models: the basis for knowledge integration and exchange, *Microbial Risk Analysis* (2018), doi: 10.1016/j.mran.2018.06.001

Highlights

- Structuring of relevant terms in microbiological risk assessment modelling domain
- Definition of metadata and controlled vocabularies for annotation of QMRA models
- Harmonized model annotation will support knowledge exchange between software tools
- Establishment of curated food safety knowledge repository is proposed
- Need of rules for knowledge annotation essential for knowledge reusability

Title

Harmonized terms, concepts and metadata for microbiological risk assessment models: the basis for knowledge integration and exchange

Authorship

Leticia Ungaretti Haberbeck*, Carolina Plaza Rodríguez*, Virginie Desvignes, Paw Dalgaard, Moez Sanaa, Laurent Guillier, Maarten Nauta, Matthias Filter

**These authors contributed equally.*

Author names and affiliations

Leticia Ungaretti Haberbeck [a] (luha@food.dtu.dk)

Carolina Plaza Rodríguez [b] (Carolina.Plaza-Rodriguez@bfr.bund.de)

Virginie Desvignes [c] (Virginie.DESVIGNES@anses.fr)

Paw Dalgaard [a] (pada@food.dtu.dk)

Moez Sanaa [c] (Moez.SANAA@anses.fr)

Laurent Guillier [c] (Laurent.GUILLIER@anses.fr)

Maarten Nauta [a] (maana@food.dtu.dk)

Matthias Filter [b] (Matthias.Filter@bfr.bund.de)

[a] Technical University of Denmark (DTU), National Food Institute (DTU Food), Kemitorvet, Buildings 201-204 on Lyngby Campus, DK-2800 Kgs. Lyngby, Denmark

[b] German Federal Institute for Risk Assessment (BfR), Department 4 - Biological Safety, Max-Dohrn-Straße 8–10, 10589 Berlin, Germany

[c] French Agency for Food, Environmental and Occupational Health & Safety (ANSES), 14 rue Pierre et Marie Curie, 94701 Maisons-Alfort Cedex, France

Corresponding authors

Laurent Guillier; Maarten Nauta; Matthias Filter

Short Title or Running Head

Harmonization for microbiological risk assessment modelling

Abstract

In the last decades the microbial food safety community has developed a variety of valuable knowledge (e.g., mathematical models and data) and resources (e.g., databases and software tools) in the areas of quantitative microbial risk assessment (QMRA) and predictive microbiology. However, the reusability of this knowledge and the exchange of information between resources are currently difficult and time consuming. This problem has increased over time due to the lack of harmonized data format and rules for knowledge annotation. It includes the lack of a common understanding of basic terms and concepts and of a harmonized information exchange format to describe and annotate knowledge. The existence of ambiguities and inconsistencies in the use of terms and concepts in the QMRA and predictive microbial (PM) modelling necessitates a consensus on their refinement, which will allow a harmonized exchange of information within these areas. Therefore, this work aims to harmonize terms and concepts used in QMRA and PM modelling spanning from high level concepts as defined by *Codex Alimentarius,* Food and Agriculture Organization (FAO) and World Health Organization (WHO), up to terms generally used in statistics or data and software science. As a result, a harmonized schema for metadata that allows consistent annotation of data and models from these two domains is proposed. This metadata schema is also a key component of the Food Safety Knowledge Markup Language (FSK-ML), a harmonized format for information exchange between resources in the QMRA and PM modelling domain. This work is carried out within a research project that aims to establish a new community resource called Risk

Assessment Modelling and Knowledge Integration Platform (RAKIP). This platform will facilitate the sharing and execution of curated QMRA and PM models using the foundation of the proposed harmonized metadata schema and information exchange format. Furthermore, it will also provide access to related open source software libraries, converter tools and software-specific import and export functions that promote the adoption of FSK-ML by the microbial food safety community. In the future, these resources will hopefully promote both the knowledge reusability and the high-quality information exchange between stakeholders within the areas of QMRA and PM modelling worldwide.

Keywords

QMRA modelling; Information exchange format; Metadata schema; Controlled vocabularies; Model annotation

1. Introduction

The management of microbiological safety of food products is a significant challenge on a global scale. Different factors contribute to this, such as the increased globalization of the food sector (Ercsey-Ravasz et al., 2012); changes in food consumption patterns (Jacxsens et al., 2010) (e.g. increased demand for minimally and/or organic processed foods) and in food preparation practices (e.g. trend of requiring better quality, fresher food and more ethical food production practices (Ragaert et al., 2004); increased resistance of some microorganisms to certain interventions (e.g. through resistance against antibiotics (Lammie and Hughes, 2016) or formation of biofilms (Carpentier and Cerf, 2011); and the introduction of new food production technologies with unknown effects on the food microbiome (Fukuda, 2015; Motarjemi and Lelieveld, 2014).

To support the management of microbial food safety risk, it is important to generate new knowledge (e.g. data and mathematical models) and resources (e.g. software tools and databases), as well as to

integrate, evaluate and apply previously generated knowledge and resources (Membré and Guillou, 2016). The predictive microbiology (Koutsoumanis et al., 2016) and quantitative microbiology risk assessment (QMRA) (Membré and Guillou, 2016) communities have invested great efforts and time to develop a rich variety of this knowledge and resources (Tenenhaus-Aziza and Ellouze, 2015). However, their reusability and the information exchange between the resources may currently be difficult and time consuming (Plaza-Rodríguez et al., 2017). This situation represents an obstacle for the performance of risk assessment using the most up to date domain knowledge. Today several predictive microbial (PM) software can export their predictions and these can then, typically with some manual data handling, be used in QMRA software tool like FDA-iRisk® (URL: https://irisk.foodrisk.org/). This is for example relevant to assess safety of new or modified foods where PM software can predict microbial responses based on essential model input including product characteristics, storage conditions and their variability (EC, 2005; Ross and Dalgaard, 2004). Clearly, it is important to improve the transfer of predictions and models from PM to QMRA software as this will timely facilitate risk assessment.

A consensus on definitions of QMRA and PM modelling terms, is an indispensable first step in achieving efficient information exchange, and will enhance transparency and confidence in the shared knowledge (Plaza-Rodríguez et al., 2017). The second step would be the creation of a conceptual framework to describe all necessary metadata on a piece of knowledge. Etymologically, metadata is "data about data". To facilitate efficient exchange of data, QMRA and PM models, it is crucial to provide a structured set of metadata that can be used by modellers or data providers to annotate their model or data set (Plaza-Rodríguez et al., 2015). The metadata will define the extent to which existing knowledge can be re-used in a transparent way. Allowing others, for instance, to apply the model in the valid range of applicability and to correctly interpret the model based prediction results. It further allows searching and finding relevant models in model repositories,

based on key words and filter functions. A conceptual metadata framework can be represented in the form of a metadata schema that may require a list of controlled vocabularies for certain metadata concepts (Sowa, 2000).

This paper follows the metadata framework and aims to to provide the foundation for the harmonization of key terms and concepts and to establish a conceptual framework for consistent annotation of knowledge from the QMRA and PM modelling domain. The harmonized terms are important for the creation of an open information exchange format what we call Food Safety Knowledge Markup Language (FSK-ML).

2. Methods

The definition of relevant terms, concepts and metadata for the QMRA and PM modelling domains were achieved through a series of workshops and web meetings carried out over the course of the RAKIP (Risk Assessment Modelling and Knowledge Integration Platform) project that has been initiated and carried out by three European risk assessment agencies: ANSES, BfR and DTU Food. The activity was structured into three main phases. First, fundamental terms and concepts relevant in the process of modelling and risk assessment were identified and defined. Second, a list of metadata concepts were collected, mapped and structured into a coherent metadata schema. Finally each metadata concept was further specified by providing it's cardinality, format, description and, if applicable, a controlled vocabulary.

2.1 Harmonization of terms and concepts

Terms and concepts describing the steps and entities in the risk assessment model generation process including PM and QMRA modelling were discussed until consensus was reached, then they were detailed in several schemata and in a glossary. In order to exploit synergies with previous work and to avoid duplicities the international recognized reports and guidelines from *Codex*

Alimentarius, Food and Agriculture Organization (FAO) and World Health Organization (WHO) (Codex Alimentarius Commission, 1999; FAO/WHO, 2016, 2009, 2008, 2003) were used if possible. In the present paper, the terms defined in the glossary and/or depicted in one of the schemata are shown in bold and italic fonts the first time they appear in the text from this point.

2.2 Establishment of a coherent metadata schema

On the basis of a common understanding of the high-level terms and concepts, a comprehensive list of metadata concepts was created that allows to annotate with sufficient detail any type of model or ***data*** in the QMRA or PM modelling domain. To accomplish this, at first, existing software tools like FDA-iRISK®, ICRA (Interactive online Catalogue on Risk Assessment; URL: http://icra.foodrisk.org/) and PMM-Lab (Predictive Microbial Modelling Lab, URL: https://foodrisklabs.bfr.bund.de/pmm-lab/) were analysed on the ***metadata*** required to annotate QMRA or PM data or models. In addition, other relevant sources like Dublin Core Metadata Initiative (URL: http://dublincore.org/), RIS format specifications (URL: goo.gl/AuPxQU) and vCard data format (URL: https://tools.ietf.org/html/rfc6350) were also explored. The collection of metadata concepts were then mapped, extended and structured into a consistent hierarchical metadata schema. The complete set of metadata concepts will be referred to as the "***Generic Metadata Schema***". Sub-sets of this schema were then generated for data and the different ***model classes***, i.e. ***predictive microbial model***, ***process model***, ***exposure model***, ***dose-response model***, ***health metric model, consumption model***, other ***empirical model***, ***risk characterization model*** and ***QMRA model***.

2.3 Detailing metadata concepts

Finally, for each metadata concept further details were specified, as its ***cardinality***, data type, description and if applicable a controlled vocabulary. By provisioning the cardinality we implicitly

defined which metadata concept was considered as mandatory for the annotation of data and the different model classes. The selection of controlled vocabularies for relevant metadata concepts was guided by the motivation to re-use as many resources as possible that are already used by the corresponding scientific community.

3. Results

3.1.Terms and concepts (phase 1)

Figure 1 depicts the QMRA ***modelling process*** within the full ***risk analysis*** process. A full risk analysis process integrates three components: ***risk assessment*** (e.g. QMRA), ***risk management***, and ***risk communication*** (Codex Alimentarius Commission, 1999). A QMRA should be initiated in response to a well-defined ***risk question*** formulated by risk managers. The risk question determines the QMRA´s objective by defining the ***hazard***, the ***product*** or ***matrix***, the relevant steps (***exposure scenario***) and the target population (***population group***); all forming the base for the construction of the QMRA ***model structure*** (Figure 1). The context, complexity and purpose of the risk question define the scope, which may or may not involve a full risk assessment, the methods and relevant tools (Dennis et al., 2008).

Based on the scope and available ***background knowledge*** the risk assessor decides which risk assessment elements can be supported by ***mathematical model***s, e.g. a dose-response model or an exposure model. Once the model structure is defined, the risk assessor has to define the ***model equations*** and ***parameters*** for the different model components and for this purpose, already existing background knowledge and newly generated data (***data collection***) can be utilized. In an iterative process the suitability of the selected model structure is evaluated against the available data and background knowledge. Alternative model equations and/or data that better fits the requirements of the risk question may be explored and chosen. Besides answering the risk question, the output of a

QMRA model can be used to identify data gaps, assess the probability of ***risk*** of an adverse health outcome and assess the effects of interventions, or in a broader context, alternative scenarios (Nauta, 2008).

Risk management at the governmental level is the process of weighing policy alternatives, in consultation with all interested parties, considering the risk assessment results and other factors relevant for the health protection of consumers and for the promotion of fair trade practices, and, if needed, selecting appropriate prevention and control options (FAO/WHO, 2016). At the food industry level, risk management includes a series of generic requirements that are included in prerequisite programmes, such as Good Manufacturing Practices or Good Hygienic Practices, GHP. Furthermore, food safety requirements that are specific for an individual processing facility and for a specific food product are included in Hazard Analysis Critical Control Points (HACCP) systems (Buchanan and Whiting, 1998; Doménech et al., 2008; Ryder et al., 2014).Risk communication is the interactive exchange of information and opinions concerning hazards, risks, risk-related factors and risk perceptions, among risk assessors, risk managers, consumers, industry, the academic community and other interested parties, including the explanation of risk assessment findings and the basis of risk management decisions (FAO/WHO, 2016). In the context of a full risk analysis, the focus of the present work is on the ***model generation process*** within the risk assessment component.

Model generation process

The term model generation process refers to a process aiming at the description of a system through mathematical concepts and/or software. The model generation process is part of the general modelling process that involves the model generation and its application. Figure 2 illustrates the different steps within the model generation process for QMRA, PM and other empirical model. Background knowledge, the foundation for the model generation process, can be extracted from

different sources as detailed in Table 1 and in Bellet et al. (2012). Background knowledge can be classified into different categories being the ***data-driven knowledge*** (e.g. predictive microbiology data, food consumption data) and ***model-based knowledge*** (e.g. PM models, models from previous risk assessments).

Generation and application of PM and empirical models

PM models can be based on common underlying mechanisms for example thermodynamic constants as suggested for the effect of temperature on growth rates (Corkrey et al., 2016; Ratkowsky et al., 2005). However, most PM models are of a more empirical nature and attempt to describe kinetic responses using relatively simple and biologically interpretable equations and parameters (Ross and Dalgaard, 2004).

For PM and empirical models the first step on the model generation process usually exploits data collected in ***experimental or observational studies*** (Figure 2). These data can be used either as ***training data*** to generate the ***fitted model***, or as ***evaluation data***, used in a ***validation procedure***. Usually, the ***empirical model generation process*** starts with a model equation obtained from background knowledge. This model equation is fitted (parameterized) to best describe the training data through a ***fitting procedure*** resulting in a fitted model. Validation of a fitted model with independent evaluation data results in a ***validated-fitted model*** when the evaluation is successful based on accepted criteria of model performance. A graphically illustrated simplified example of the model generation process of a PM model is available in the supplementary material (Figure S1). Generation and validation of PM models are an important and a demanding activity where some studies used several hundred kinetic curves to create and validate a PM model and it has been recommended to use indices of model performance including bias factor, accuracy factor and the acceptable simulation zone approach in combination with graphical methods (Mejlholm et al., 2010; Østergaard et al., 2014). Examples of other empirical models are time-temperature models. These

models aims to predict temperature profiles along the food chain according to operational conditions like external temperature and thermostat setting (Laguerre et al., 2014; Laguerre and Flick, 2010; Lecoq et al., 2016). The generation of dose-response model***s*** or epidemiological models also follows the process.

The knowledge generated through empirical and PM models is frequently highly relevant in QMRA studies. The application of validated-fitted PM models, for instance, can help to understand the impact of unit operations along the food production chain on the number of microorganism per product provided that the product characteristics and variability are determined and that the storage conditions are realistic (Lammerding and McKellar, 2004; Nauta, 2002).

QMRA models

The model generation process for QMRA models is usually developed along the four well-established elements of a risk assessment as defined by the *Codex Alimentarius* Commission: (i) ***hazard identification***, (ii) ***hazard characterization***, (iii) ***exposure assessment*** and (iv) ***risk characterization*** (Codex Alimentarius Commission, 1999). Figure 3 depicts these four high level elements and the different model classes that belong to them, such as dose-response model, process model, consumption model, exposure model, health metric model and risk characterization model. The model classes can be combined to form a full QMRA model.

Hazard identification is the identification of biological, chemical and physical agents (Hazard) capable of causing adverse health effects and which may be present in a particular group of food product (Product/Matrix) where the hazard may be present (FAO/WHO, 2016). The present work focuses exclusively on the ***biological hazards***. Population group refers to the characterization of the different consumer groups or target groups that could be categorized by age, gender, ethnic origin, health status, culture/region, socio-economic factors, etc. (FAO/WHO, 2008). These three elements

are normally determined in the risk question agreed between the risk managers and risk assessors (Nauta, 2008).

Hazard characterization provides a qualitative and/or quantitative evaluation of the nature of the adverse health effects that may result from the ingestion of a hazard which may be present in food or feed products (FAO/WHO, 2016). This step might include a dose-response model if available. If there are no means to define a credible dose-response relationship or to determine the level of exposure, an alternative approach has to be found (FAO/WHO, 2009).

Exposure assessment is the qualitative and/or quantitative evaluation of the likely intake of the hazard via the feed or food product (Product/Matrix) as well as exposures from other sources if relevant (FAO/WHO, 2016). The exposure assessment usually contains a process model and a consumption model. The combination of these two models forms an exposure model. The process model describes the "farm-to-fork" food production chain (Fig. 3). Several approaches have been suggested for such models, we suggest the use of the modular process risk models (MPRMs), proposed by Nauta (2001). The idea behind MPRMs is that the process model can be divided into ***modules***. And, in principle, each module is defined to reflect one of six basic processes: growth, inactivation, mixing, partitioning, removal and cross-contamination (Nauta, 2008). This principle is also used in FDA-iRISK®, be it with a slightly different categorization of basic processes/process types. A module may combine several processing steps if they have a similar impact on the microorganism, such as in Van Damme et al. (2017), where cutting and deriding of pork bellies produce the same basic process (removal) and thus were combined in a single module.

The last element of a QMRA is the risk characterization (Figure 3). This element comprises the qualitative and/or quantitative estimation, including attendant uncertainties, of the probability of occurrence and severity of known or potential adverse health effects in a given population based on the results of the previous steps of hazard identification, hazard characterization and exposure

assessment (FAO/WHO, 2016). Within the framework of the risk characterization element, the combination of health metrics model, dose-response model and exposure model results in a risk characterization model. The combination of all model classes will form a full QMRA model (Fig. 3). Graphically illustrated Figure S2 (Supplementary material) depicts an example of the composition of QMRA elements displayed in Figure 3.

Once the QMRA model has been generated, a ***reviewing procedure*** needs to be applied in order to be able to use it for ***decision-making*** purposes by risk managers. The reviewing process aims to assess the quality of a risk assessment, checking all the assumptions, technicalities, logic of the model, data and model selection, etc. It can be accomplished as a multistep process involving one or more technical reviews, regulatory reviews and independent formal peer review (Dennis et al., 2008).

To reach a comprehensive coverage of relevant terms and concepts we established an online glossary that can be freely accessed and updated (URL: goo.gl/b4ADho) by the community. So far, the ***RAKIP*** partners are the curators of this online glossary.

3.2. Metadata schema for knowledge annotation (phases 2 and 3)

Based on the description of general terms and concepts, we introduce in this section a comprehensive schema of metadata concepts for model and data annotation in the QMRA area. This schema is named "Generic Metadata Schema" (supplementary material, Table S1). In there the metadata concepts are organized in a hierarchical structure consisting of three levels. Level 1 ("Top level") defines four information areas on which metadata are collected: (i) "General information", (ii) "Scope", (iii) "Data background" and (iv) "Model math / Data definition" (first column in Table S1). Level 2 ("Topic") provides for each level 1 area (e.g. "Scope") an exhaustive list of metadata concepts where each might further be detailed through level 3 ("Detailed metadata concept") entries

(see third and fifth column in Table S1). The "Generic Metadata Schema" contains in addition a cardinality value (1, 0:1, 0:N, 1:N) for each metadata concept on each level. This cardinality value indicates if the metadata concept is mandatory or not and its possible ***dimension***. The value of "1" means that the concept is mandatory, i.e. a metadata necessary for providing the key elements for the usability and accessibility of data/models. The value of "0:1" is used if metadata are not mandatory, but in case information is provided, just one is allowed. The values of "0:N" and "1:N" allow several entries; while in the first case it is not mandatory to provide information, in the second one it is.

In a second step, the "Generic Metadata Schema" was used as a template for the generation of specific metadata schema for data and model classes (Table S2). The underlying idea here is to maintain the highest possible level of similarity between the domain specific metadata schema, whilst providing necessary flexibility in metadata naming and pre-selection. As part of the adjustment also the cardinality property could be adapted and even changed to the value of "X" means that the metadata is not relevant to describe the model class. For instance, the metadata Product/Matrix is mandatory for the model class PM model, process model, consumption model, risk characterization model and QMRA model, while for other model classes it is not mandatory or irrelevant (Table S2).

Finally, a list of controlled vocabularies was provided for as many metadata concepts as possible using existing resources like public controlled vocabularies, information standards and software tools (Table S1). Most of the controlled vocabularies were taken from the Standard Sample Description for Food and Feed ver. 2.0 (SSD) created by the European Food Safety Agency (EFSA) (EFSA, 2013). This standard was created to facilitate reporting of surveillance data to EFSA from several food safety domains. As it is broadly accepted, it is a valuable source for the metadata schema. For example, we attribute a controlled vocabulary from SSD to annotate the metadata

Product/Matrix. A total of 386 terms are available to describe precisely this metadata concept according to its origin such as animal or vegetal among others. In addition some vocabularies used in domain-specific software tools like PMM-Lab, ICRA and FDA-iRisk®, were also included. Some terms for specific metadata have been taken from other ontologies or public standards. For example for the Publication type/status metadata, we use the Bibliographic Ontology Specifications (URL: http://bibliontology.com/); for metadata on parameter distributions the probONTO (Swat et al., 2016) ontology and for naming populations the FOODON (URL: https://foodontology.github.io/foodon/) resource.

The complete metadata schema is hosted online in the so called "Metadata Master Table" (URL: https://goo.gl/PE4ysP) and the controlled vocabularies are made available online (URL: https://goo.gl/wbFoZU).

3.3. Food Safety Knowledge Markup Language (FSK-ML)

The terms, concepts and metadata schema presented are an important foundation for the creation of the first specification for the Food Safety Knowledge Markup Language (***FSK-ML***) (URL: https://foodrisklabs.bfr.bund.de/fsk-ml-food-safety-knowledge-markup-language/). The FSK-ML specification is a software developer guidance document that explicitly specifies the structure and content of the files (***FSKX-file***) that may be used to encode models in the domain of food safety risk assessment in the future. Similar markup languages are used in different scientific domains like the Systems Biology Markup Language (SBML) (URL: http://sbml.org/Main_Page). The development of FSK-ML is crucial to enable sharing of data, QMRA, empirical models and PM models between different software tools and model repositories. With this information exchange format it will also be guaranteed, that metadata stay linked to the correct model or data set. A specific feature of FSK-ML is that it supports the exchange of models that are provided in specific script-based programming languages (e.g. R, Matlab and Python).

4. Discussion

The harmonized terms, concepts and metadata presented in this paper are an important first result of the RAKIP project, a collaboration between three European food safety risk assessment institutions (ANSES, BfR and DTU Food). They were extensively discussed by the RAKIP partners, with the objective to create the foundation for the FSK-ML format. They are a requirement to create new community resources that support knowledge integration and exchange within the PM and QMRA areas, which is the main objective of the project. Besides the creation of the FSK-ML language, RAKIP aims at the creation of a web portal (URL: https://foodrisklabs.bfr.bund.de/rakip-web-portal/) that allows the microbiological food safety community to (i) contribute to the definition on harmonized terms and concepts (ii) share software libraries, converter tools and software-specific import and export functions promoting the adoption of FSK-ML by the risk assessment community; (iii) share FSK-ML compliant models and data in a model repository with search and filtering functionalities that also allows the download, execution and modification of models.

Continuous effort is needed to maintain and update the resources presented in this paper, to assure and to create a broad compliance and support within the food safety community (including food authorities, food industries, consultancy companies and food research institutes). Currently the RAKIP project focussed primarily on terms, concepts and metadata from the areas of QMRA and PM modelling. However in the future, other areas, like chemical risk assessment, and other modelling approaches, as e.g. Bayesian networks (Beaudequin et al., 2015) or machine learning (Laabei et al., 2014), should be addressed. As this research has been initiated as an open community effort it can take up suggestions on future development goals from the scientific community. In addition, the RAKIP project embraces the idea of “Open Source” and “Open Access”, and wherever possible it intends to generate synergies with other existing projects and initiatives.

Currently, the RAKIP project main focus is to provide existing QMRA or PM models in the proposed FSK-ML format which serves as proof-of-principle for the underlying knowledge exchange concept. Besides that, the proof-of-principle will lead to improvements in the current metadata scheme. This work will also support the development of a "Minimal Information Required in the Annotation of Risk Assessment Models" (MIRARAM) guideline. This guideline would be similar to the MIRIAM guideline which is widely accepted in the Systems Biology community (Novere et al., 2005), and would provide a set of rules for knowledge annotation, that would be used in the future as a basic for the establishment of a ***curation*** process for those models intended to be shared in the repository.

This and related work carried out by RAKIP partners (e.g. the development of the open-source FSK-Lab software and the new ***RAKIP web portal***) will support different ***end user*** within the microbiological food safety community. For example: (a) modellers would be able to share their models in a harmonized way with the scientific community, (b) reviewers of research papers exploiting these models would be able to easily reproduce results from the paper, (c) risk assessors could easier exploit the scientific knowledge generated by the research community by having QMRA and EM models readily available with a harmonized description which consequently would allow risk assessors to perform their work in a shorter time. More details on the end users objectives and on the description of end users cases are given by Plaza-Rodríguez et al. (2017).

5. Conclusion

The *Codex Alimentarius* commission recommends that the risk analysis process remains open, fully and systematically documented in a transparent manner. In addition, it should be evaluated and reviewed as appropriate in the light of newly generated scientific data by the food safety and science communities (FAO/WHO, 2016). The present work proposes a conceptual framework on terms, concepts and metadata that could serve as a foundation for harmonized annotation of risk

assessment models and linked data. The adoption of such a framework within the risk assessment modelling domain will contribute to enhancing transparency and will also facilitate the exchange of data and models between different software tools thanks to the development of the FSK-ML standard. In this sense, the RAKIP project supports the establishment of resources for sharing and re-using of knowledge in a transparent way, e.g. through harmonized information exchange formats, rules for model annotation and a web platform dedicated to microbial food safety risk assessment community. Ultimately, this will facilitate faster high quality risk assessment and decision-making for food safety managers that follows harmonized international standards.

Acknowledgements

We thank Miguel de Alba Aparicio, Taras Günther, Lars Valentin, Octavio Mesa Varona and Ahmad Swaid for their invaluable contribution on the creation of the RAKIP web portal demonstrator and all the other technological resources that have been necessary to make this project a reality.

Funding

This work was jointly supported by the German Federal Institute for Risk Assessment (BfR), the National Food Institute (DTU Food) from the Technical University of Denmark (DTU) and the French Agency for Food, Environmental and Occupational Health & Safety (ANSES). This work has been supported in part by the AGINFRA PLUS project that is funded by the European Commission's Horizon 2020 research and innovation programme under grant agreement No 731001, the EFSA-BfR Framework Partnership Agreement GP/EFSA/AMU/2016/01 and by funds of the Federal Ministry of Food and Agriculture (BMEL) for the project "FoodAuthent" based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support programme. The content of this paper

does not reflect the official opinion of the European Union, EFSA, BMEL or BLE. Responsibilityfor the information and views expressed in this paper lies entirely with the author(s).

References

Beaudequin, D., Harden, F., Roiko, A., Stratton, H., Lemckert, C., Mengersen, K., 2015. Beyond QMRA: Modelling microbial health risk as a complex system using Bayesian networks. Environ. Int. 80, 8–18. doi:https://doi.org/10.1016/j.envint.2015.03.013

Bellet, C., Humblet, M., Swanenburg, M., Dhé, J., Vandeputte, S., Thébault, A., Gauchard, F., Hendrikx, P., De Vos, C., De Koeijer, A., Saegerman, C., Sanaa, M., 2012. Specification of data collection on animal diseases to increase the preparedness of the AHAW panel to answer future mandates – CFP/EFSA/AHAW/2010/01. EFSA Support. Publ. 9, EN-354. doi:10.2903/sp.efsa.2012.EN-354

Buchanan, R.L., Whiting, R.C., 1998. Risk assessment: a means for linking HACCP plans and public health. J. Food Prot. 61, 1531–1534.

Buchanan, R.L., Whiting, R.C., 1997. Concepts in Predictive Microbiology. 50th Annu. Reciprocal Meat Conf. 50, 93–97.

Carpentier, B., Cerf, O., 2011. Review - Persistence of *Listeria monocytogenes* in food industry equipment and premises. Int. J. Food Microbiol. 145, 1–8. doi:10.1016/j.ijfoodmicro.2011.01.005

Codex Alimentarius Commission, 2014. Principles and Guidelines for the Conduct of Microbiological Risk Assessment. CAC/GL 30-1999. Adopted 1999. Amendments 2012, 2014.

Codex Alimentarius Commission, 1999. Principles and Guidelines for the Conduct of Microbiological Risk Assessment, in: Joint FAO/WHO Food Standards Programme, (Ed.),

CAC/GL-30, Rome.

Corkrey, R., McMeekin, T.A., Bowman, J.P., Ratkowsky, D.A., Olley, J., Ross, T., 2016. The Biokinetic Spectrum for Temperature. PLoS One 11, e0153343. doi:10.1371/journal.pone.0153343

Dennis, S.B., Kause, K., Losikoff, M., Engeljohn, D.L., Buchanan, R.L., 2008. Using Risk Analysis for Microbial Food Safety Regulatory Decision Making, in: Schaffner, D.W. (Ed.), Microbial Risk Analysis of Foods. ASM Press, Washington, D.C., pp. 137–175.

Doménech, E., Escriche, I., Martorell, S., 2008. Assessing the effectiveness of critical control points to guarantee food safety. Food Control 19, 557–565. doi:10.1016/j.foodcont.2007.06.015

EC, 2005. Commission regulation (EC) No 2073/2005 of 15 November 2005 on microbiological criteria for foodstuffs. Off. J. Eur. Union 338, 1–25.

EFSA, 2013. Standard Sample Description ver. 2.0. EFSA J. 11, 3424. doi:10.2903/j.efsa.2013.3424

Ercsey-Ravasz, M., Toroczkai, Z., Lakner, Z., Baranyi, J., 2012. Complexity of the International Agro-Food Trade Network and Its Impact on Food Safety. PLoS One 7, e37810. doi:10.1371/journal.pone.0037810

FAO/WHO, 2016. Codex Alimentarius Commission. Procedural manual, 25th ed. Rome.

FAO/WHO, 2009. Risk Characterization of Microbiological Hazards in Food: Guidelines. Microbiol. Risk Assess. Ser. No. 17. 135 p. doi:http://dx.doi.org/10.1111/j.1539-6924.2012.01825.x

FAO/WHO, 2008. Exposure assessment of microbiological hazards in food: Guidelines. Microbiol. Risk Assess. Ser. No. 7. 61, 102 p.

FAO/WHO, 2003. Hazard Characterization for Pathogens in Food and Water: Guidelines.

Microbiol. Risk Assess. Ser. No. 3. 61 p.

Fukuda, K., 2015. Food safety in a globalized world. Bull. World Heal. Organ 93, 212. doi:http://dx.doi.org/10.2471/BLT.15.154831

ISO, 2004. (International Organisation for Standardisation). Information technology -- Metadata registries (MDR) Part 1: Fr ISO/IEC 11179-1.

Jacxsens, L., Luning, P.A., van der Vorst, J.G.A.J., Devlieghere, F., Leemans, R., Uyttendaele, M., 2010. Simulation modelling and risk assessment as tools to identify the impact of climate change on microbiological food safety – The case study of fresh produce supply chain. Food Res. Int. 43, 1925–1935. doi:https://doi.org/10.1016/j.foodres.2009.07.009

Koutsoumanis, K.P., Lianou, A., Gougouli, M., 2016. Latest developments in foodborne pathogens modeling. Curr. Opin. Food Sci. 8, 89–98. doi:10.1016/j.cofs.2016.04.006

Laabei, M., Recker, M., Rudkin, J.K., Aldeljawi, M., Gulay, Z., Sloan, T.J., Williams, P., Endres, J.L., Bayles, K.W., Fey, P.D., Yajjala, V.K., Widhelm, T., Hawkins, E., Lewis, K., Parfett, S., Scowen, L., Peacock, S.J., Holden, M., Wilson, D., Read, T.D., Van Den Elsen, J., Priest, N.K., Feil, E.J., Hurst, L.D., Josefsson, E., Massey, R.C., 2014. Predicting the virulence of MRSA from its genome sequence. Genome Res. 24, 839–849. doi:10.1101/gr.165415.113

Laguerre, O., Duret, S., Hoang, H.M., Flick, D., 2014. Using simplified models of cold chain equipment to assess the influence of operating conditions and equipment design on cold chain performance. Int. J. Refrig. 47, 120–133. doi:https://doi.org/10.1016/j.ijrefrig.2014.07.023

Laguerre, O., Flick, D., 2010. Temperature prediction in domestic refrigerators: Deterministic and stochastic approaches. Int. J. Refrig. 33, 41–51. doi:https://doi.org/10.1016/j.ijrefrig.2009.09.014

Lammerding, A.M., McKellar, R.C., 2004. Predictive Microbiology in Quantitative Risk

Assessment, in: McKellar, R.C., Lu, X. (Eds.), Modeling Microbial Responses in Foods. CRC Press, Boca Raton, Florida, pp. 274–295.

Lammie, S.L., Hughes, J.M., 2016. Antimicrobial Resistance, Food Safety, and One Health: The Need for Convergence. Annu. Rev. Food Sci. Technol. 7, 287–312. doi:10.1146/annurev-food-041715-033251

Lecoq, L., Flick, D., Derens, E., Hoang, H.M., Laguerre, O., 2016. Simplified heat and mass transfer modeling in a food processing plant. J. Food Eng. 171, 1–13. doi:https://doi.org/10.1016/j.jfoodeng.2015.09.026

Mejlholm, O., Gunvig, A., Borggaard, C., Blom-Hanssen, J., Mellefont, L., Ross, T., Leroi, F., Else, T., Visser, D., Dalgaard, P., 2010. Predicting growth rates and growth boundary of *Listeria monocytogenes* - An international validation study with focus on processed and ready-to-eat meat and seafood. Int. J. Food Microbiol. 141, 137–50. doi:10.1016/j.ijfoodmicro.2010.04.026

Membré, J.M., Guillou, S., 2016. Lastest developments in foodborne pathogen risk assessment. Curr. Opin. Food Sci. 8, 120–126. doi:10.1016/j.cofs.2016.04.011

Motarjemi, Y., Lelieveld, H., 2014. Fundamentals in Management of Food Safety in the Industrial Setting: Challenges and Outlook of the 21st Century, in: Motarjemi, Y., Lelieveld, H. (Eds.), Food Safety Management. A Practical Guide for the Food Industry. Elsevier, pp. 1–20. doi:http://dx.doi.org/10.1016/B978-0-12-381504-0.00001-9

Nauta, M.J., 2008. The Modular Process Risk Model (MPRM): a Structured Approach to Food Chain Exposure Assessment, in: Schaffner, D.W. (Ed.), Microbial Risk Analysis of Foods. ASM Press, Washington, D.C., pp. 99–136.

Nauta, M.J., 2002. Modelling bacterial growth in quantitative microbiological risk assessment: Is it

possible? Int. J. Food Microbiol. 73, 297–304. doi:10.1016/S0168-1605(01)00664-X

Nauta, M.J., 2001. A modular process risk model structure for quantitative microbiological risk assessment and its application in an exposure assessment of *Bacillus cereus* in a REPFED. report nr. 149106 007 RIVM. Bilthoven, The Netherlands.

Novere, N. Le, Finney, A., Hucka, M., Bhalla, U.S., Campagne, F., Collado-Vides, J., Crampin, E.J., Halstead, M., Klipp, E., Mendes, P., Nielsen, P., Sauro, H., Shapiro, B., Snoep, J.L., Spence, H.D., Wanner, B.L., 2005. Minimum information requested in the annotation of biochemical models (MIRIAM). Nat Biotech 23, 1509–1515.

Østergaard, N.B., Eklöw, A., Dalgaard, P., 2014. Modelling the effect of lactic acid bacteria from starter- and aroma culture on growth of *Listeria monocytogenes* in cottage cheese. Int. J. Food Microbiol. 188, 15–25. doi:10.1016/j.ijfoodmicro.2014.07.012

Plaza-Rodríguez, C., Haberbeck, L.U., Desvignes, V., Dalgaard, P., Sanaa, M., Nauta, M., Filter, M., Guillier, L., 2017. Towards transparent and consistent exchange of knowledge for improved microbiological food safety. Curr. Opin. Food Sci. In press. doi:10.1016/j.cofs.2017.12.002

Plaza-Rodríguez, C., Thoens, C., Falenski, A., Weiser, A.A., Appel, B., Kaesbohrer, A., Filter, M., 2015. A strategy to establish Food Safety Model Repositories. Int. J. Food Microbiol. 204, 81–90. doi:10.1016/j.ijfoodmicro.2015.03.010

Ragaert, P., Verbeke, W., Devlieghere, F., Debevere, J., 2004. Consumer perception and choice of minimally processed vegetables and packaged fruits. Food Qual. Prefer. 15, 259–270. doi:10.1016/S0950-3293(03)00066-1

Ratkowsky, D.A., Olley, J., McMeekin, T.A., Ball, A., 1982. Relationship Between Temperature and Growth Rate of Bacterial Cultures. J. Bacteriol. 149, 1–5.

Ratkowsky, D.A., Olley, J., Ross, T., 2005. Unifying temperature effects on the growth rate of bacteria and the stability of globular proteins. J. Theor. Biol. 233, 351–362. doi:https://doi.org/10.1016/j.jtbi.2004.10.016

Ross, T., Dalgaard, P., 2004. Secondary models, in: McKellar, R.C., Lu, X. (Eds.), Modeling Microbial Responses in Foods. CRC Press LLC, NY, USA, pp. 64–122.

Rosso, L., Lobry, J.R., Flandrois, J.P., 1993. An Unexpected Correlation between Cardinal Temperatures of Microbial Growth Highlighted by a New Model. J. Theor. Biol. 162, 447–463.

Ryder, J., Iddya, K., Ababouch, L., 2014. Assessment and management of seafood safety and quality - Current practices and emerging issues. FAO Fisheries and Aquaculture Technical Paper No. 574. Rome, 432 pp.

Sowa, J.F., 2000. Ontology, Metadata, and Semiotics BT - Conceptual Structures: Logical, Linguistic, and Computational Issues, in: Ganter, B., Mineau, G.W. (Eds.), . Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 55–81.

Swat, M.J., Grenon, P., Wimalaratne, S., 2016. ProbOnto: Ontology and knowledge base of probability distributions. Bioinformatics 32, 2719–2721. doi:10.1093/bioinformatics/btw170

Tenenhaus-Aziza, F., Ellouze, M., 2015. Software for predictive microbiology and risk assessment: A description and comparison of tools presented at the ICPMF8 Software Fair. Food Microbiol. 45, 290–299. doi:10.1016/j.fm.2014.06.026

Van Damme, I., De Zutter, L., Jacxsens, L., Nauta, M.J., 2017. Control of human pathogenic *Yersinia enterocolitica* in minced meat: Comparative analysis of different interventions using a risk assessment approach. Food Microbiol. 64, 83–95. doi:10.1016/j.fm.2016.12.006

Whiting, R.C., Buchanan, R.L., 1993. A classification of models in predictive microbiology - a

reply to K.R. Davey. Food Microbiol. 10, 175–177.

Glossary

The glossary contains the definition of the most important terms and concepts presented in this manuscript. An extended version containing other relevant terms for the RAKIP project is also online (URL: goo.gl/b4ADho), where the definitions can be updated by the community and further curated by the RAKIP partners. The following definitions were last updated on February 19^{h} 2018. The terms not containing a reference were agreed by the authors

- *Background knowledge*: Is the foundation of the model generation process. Examples of background knowledge sources in QMRA area are: published scientific studies, unpublished studies, national or international official data, surveys and questionnaires, risk assessments carried out by national or international agencies and research institutes, expert knowledge elicitation and expert opinion and events (e.g. meeting or conference). It can be classified, for instance, into data-driven knowledge (e.g. predictive microbiology data, food consumption data) and model-based knowledge (e.g. PM models, models from previous risk assessments).
- *Biological hazard*: The microorganisms and/or their toxins present in, or condition of, food with the potential to cause an adverse health effect (adapted from *Hazard* definition from FAO/WHO, 2016).
- *Cardinality*: The cardinality value indicates if the metadata concept is mandatory or not, and the possible dimension. The value of "1" means that the metadata is mandatory. The value of "0:1" means that the metadata concepts are not mandatory, but in case information is provided, just one is allowed. The values of "0:N" and "1:N" allow several entries, where in the first case it is not mandatory to provide information, in the second one it is. "X" means that the metadata is not relevant to describe the model class.
- *Consumption model*: A consumption model describes the amount of food consumed during a particular eating occasion (i.e., a serving) and/or the frequency of the consumption of these

servings, or an average amount of food consumed per day. This amount may vary in time, between individuals, betweenthe different population groups of interest and the considered exposure type.

- *Curation*: A curation process for models aims to classify models according to certain quality criteria. These quality criteria must be defined. Depending on the classification, the end users of the model can e.g. judge how reliable the model might be.
- *Data*: Symbolic representation of observable properties of the world.
- *Data collection*: The process of gathering data. Different types of data can be collected (microbial concentrations, daily consumption, etc.). The collected data should be fit for purpose, representative and allow a meaningful analysis.
- *Data-driven knowledge*: Knowledge derived from experimental/observational data.
- *Decision-making:* The action or process of making decisions based on scientific facts or results. In the context of risk management, decision-making considers the risk assessment results and other factors relevant for the health protection of consumers and for the promotion of fair trade practices.
- *Dimension:* Possible number of elements constituting a metadata concept. See *Cardinality* for more details.
- *Dose-response model*: Model describing the "relationship between the magnitude of exposure (dose) to a hazard and the severity and/or frequency of associated adverse effects (response)" (Codex Alimentarius Commission, 1999; FAO/WHO, 2016).
- *Empirical model*: Empirical models describe a set of data in a convenient mathematical relationship without considering any underlying phenomena (Ross and Dalgaard 2004). Examples of empirical models used in the microbial food safety area: the secondary predictive microbial models of square-root-type or Ratkowsky-type (Ratkowsky et al., 1982) and cardinal

parameter models (Rosso et al., 1993), the log-logistic, log-probit and Weibull(-Gamma) models as dose-response models (FAO/WHO, 2003), time-temperature models (Laguerre et al., 2014; Laguerre and Flick, 2010; Lecoq et al., 2016) and epidemiological models.

- *Empirical model generation process*: Is founded on the relationship between observed data and approximate representations of the real systems that generated that data.
- *End user*: Person who ultimately applies or is intended to ultimately apply a model. In this aspect several levels of end users can be defined:
    - *Non-expert user*: Person who intendeds to use the model mainly for making predictions without too much interest in the modelling process.
    - *Expert user*: Person who is interested not only in the model predictions, but also in the modelling process, including data, model equation, fitting process, etc.
- *Evaluation data*: Experimental/observational data that is used to evaluate the performance of the fitted model. This data should have not been used previously for generation of the fitted model.
- *Experimental/Observational data*: Measurement values obtained in experimental or observational studies. Can be used in the fitting or validation procedure.
- *Experimental or Observational study*: In an experimental study the researcher controls the study by assigning “treatments” to a group of subjects, meanwhile in an observational study, the researchers simply “observe” a group of subjects without actually “doing" anything to the subjects. In both cases characterization of the subject (e.g. product or matrix) is essential for obtaining valuable data for model generation (training data) or evaluation (evaluation data).
- *Exposure assessment:* The qualitative and/or quantitative evaluation of the likely intake of biological, chemical, and physical agents via food as well as exposures from other sources if relevant (FAO/WHO, 2016).

- *Exposure model*: A combination of the process model and the consumption model that results in the exposure assessment.
- *Exposure scenario*: Scenarios can be constructed to predict the range of possible exposures. The scenarios might reflect effects of processing, such as hygienic design, cleaning and disinfection, as well as the time/temperature and other conditions of the food history, food handling and consumption patterns, regulatory controls, and surveillance systems (Codex Alimentarius Commission, 2014).
- *Fitted model*: Model equation that has been fitted to Experimental/Observational data (training data) with the aim of obtaining a validated-fitted model.
- *Fitting procedure*: Occurs when a model equation is fitted to Experimental/Observational data (training data) by the use of curve-fitting tools or by inference (e.g., Bayesian inference). The curve-fitting tools iteratively try different values for the model's parameters, measuring the "goodness of fit" between the model and the experimental data. This process continues until the software finds the combination of values for the model equations' parameters that provides the best fit achievable (Buchanan and Whiting, 1997).
- *FSK-ML (Food Safety Knowledge - Markup Language)*: FSK-ML is a specification for knowledge exchange format that can be used to encode models in the domain of food safety risk assessment.
- *FSKX-file*: File format defined within the FSK-ML specifications supporting the exchange of risk assessment models.
- *Generic Metadata Schema*: The complete set of metadata concepts generated as part of the RAKIP project that allows annotating with sufficient detail any type of model or data in the QMRA or PM modelling domain.

- *Hazard*: A biological, chemical or physical agent in, or condition of, food with the potential to cause an adverse health effect (FAO/WHO, 2016).
- *Hazard characterization*: The qualitative and/or quantitative evaluation of the nature of the adverse health effects associated with biological, chemical and physical agents which may be present in food (FAO/WHO, 2016).
- *Hazard identification*: The identification of biological, chemical, and physical agents capable of causing adverse health effects and which may be present in a particular food or group of foods (FAO/WHO, 2016).
- *Health metric model*: Model for calculating a measure for assessing the health impact of a specific hazard in a population group: e.g. Disability-adjusted life years (DALYs) or cost per illness.
- *Mathematical model*: A parameterized model equation (Model equation + parameters value).
- *Matrix*: The matrix for which the model or data applies. The term Matrix is inspired in the Standard Sample Description for Food and Feed ver. 2.0 (SSD) (EFSA, 2013) where the type of matrices are: food, food stimulants, feed, animal, environmental samples and food contact material.
- *Metadata*: Data that defines and describes other data (ISO, 2004).
- *Model-based knowledge*: Knowledge on a system stored inside or obtained from a model, e.g. a mathematical algorithm, EM, PM and QMRA model, software code etc.
- *Model class*: Classification of mathematical models relevant for microbial food safety such as QMRA model, dose-response model, process model, consumption model, exposure model, health metric model, predictive microbial model and other empirical model. This classification is based mainly in the risk assessment definition provided by Codex Alimentarius (Codex Alimentarius Commission, 1999).

- *Model equation*: Mathematical equation describing the relation between dependent and independent variables. See examples in "Predictive microbial model".
- *Model generation process*: Process aiming at generating a description of a system through mathematical concepts or software code.
- *Modelling process*: Processes that involve the model generation process and include the application of the generated model.
- *Model structure*: Describes the conceptual design, components and structure of a model. For instance, the QMRA model structure usually follows the well-stabilised four steps: hazard identification, hazard characterization, exposure assessment and risk characterization, while for the exposure assessment step the model structure usually follows the different phases of the food production chain.
- *Module*: The module definition comes from the Modular Process Risk Model methodology developed by Nauta (2001) where the food production chain is divided into modules. In principle, each module is defined to reflect one of the six basic process: growth, inactivation, mixing, partitioning, removal and cross-contamination (Nauta, 2001). A module may combine several processing steps if they have a similar impact on the microorganism. In FDA-iRisk® the term "process stage" is used for the same concept.
- *Parameter*: The term parameter is used in a number of ways. In the context of RAKIP project, examples of parameter are: the prevalence and concentration of foodborne pathogens on food products along the food production chain, microbial growth or inactivation rates, temperature and time profiles, cross-contamination rates, transfer rates, product characteristics (i.e. pH, water activity, concertation of organic acids) These parameters can be fixed values or variability distributions.

- *Population group*: Is the population or sub-population of interest when performing a risk assessment, for example: general population, pregnant or lactating women, seniors, infants.
- *Predictive microbial (PM) model*: Models describing the microbial responses towards environmental conditions, such as storage and processing conditions and product characteristics. Traditionally, models in predictive microbiology are classified as primary and secondary (Whiting and Buchanan, 1993):
    - *Primary models* describe the changes in microorganism concentration (e.g. during growth, survival and inactivation) according to time. The dependent variable is normally the microbial concentration and the independent variable is the time (see Model Equation definition).
    - *Secondary models* describe the parameters appearing in the primary models (e.g., the lag phase, the growth rate and inactivation rate) as a function of the environmental factors (e.g. temperature, pH, water activity, organic acids concentration). The dependent variables are the primary model parameters and the independent variables are environmental factors (see Model Equation definition).
- *Process*: A series of actions or steps taken in order to achieve a particular end. See "Fitting procedure" for an example.
- *Process model*: Model that describes how the concentrations of the hazard change along the different steps (modules) of the food production chain (potentially from farm to fork).
- *Product*: The food product for which the model or data apply.
- *QMRA model:* In the context of the model classes defined in RAKIP, QMRA model is the combination of all other model classes (see Model classes definition) and data.

- *RAKIP*: "Risk Assessment Modelling and Knowledge Platform". A project carried out jointly by ANSES, BfR and DTU in 2017.
- *RAKIP web portal*: Online resource for RAKIP partners with a Graphical User Interface for browsing, searching, filtering and downloading FSKX-files. It allows sharing FSK-ML formatted data and knowledge between the RAKIP project partners. URL: https://foodrisklabs.bfr.bund.de/rakip-web-portal/
- *Reviewing procedure*: Process aiming to access the quality of a risk assessment document, a risk assessment model, and the underlying data. Can be a multistep review process involving one or more technical reviews, regulatory reviews and independent formal peer review (Dennis et al., 2008).
- *Risk*: A function of the probability of an adverse health effect and the severity of that effect, consequential to a hazard in food (Codex Alimentarius Commission, 1999).
- *Risk analysis*: A process consisting of three components: risk assessment, risk management and risk communication (Codex Alimentarius Commission, 1999).
- *Risk assessment*: A scientifically based process consisting of the following steps: (i) hazard identification, (ii) hazard characterization, (iii) exposure assessment, and (iv) risk characterization (Codex Alimentarius Commission, 1999; FAO/WHO, 2016).
- *Risk characterization*: The qualitative and/or quantitative estimation, including attendant uncertainties, of the probability of occurrence and severity of known or potential adverse health effects in a given population based on hazard identification, hazard characterization and exposure assessment (Codex Alimentarius Commission, 1999; FAO/WHO, 2016)
- *Risk characterization model*: Combination of health metrics model, dose-response model and exposure model within the framework of the risk characterization.

- *Risk communication*: The interactive exchange of information and opinions concerning risk among risk assessors, risk managers, consumers and other interested parties (Codex Alimentarius Commission, 1999; FAO/WHO, 2016).
- *Risk management*: The process of weighing policy alternatives in the light of the results of risk assessment and, if required, selecting and implementing appropriate control options, including regulatory measures (Codex Alimentarius Commission, 1999; FAO/WHO, 2016)
- *Risk question*: Initial statement on the specific objectives of a risk assessment.
- *Training data*: Experimental/Observational data used for the generation of a fitted model.
- *Transformation*: A process by which the data is manipulated (i.e. data set is cleaned or rearranged) to fit the purpose of the model.
- *Validated-fitted model*: Model that has been fitted and successfully validated on independent experimental/observational data (evaluation data). A validated-fitted model can be used to make predictions within the range of environmental factors values where the model has been successfully validated.
- *Validation procedure*: Procedure that aims to assess the performance of fitted models and to determine if they can be used to aid decision-making. The validation process might use several hundred of evaluation data to evaluate the model accuracy. It consists of comparing model predictions with independent experimental/observational data though indices of model performance including bias factor, accuracy factor and the acceptable simulation zone approach in combination with graphical methods (Mejlholm et al., 2010; Østergaard et al., 2014).
- *Visualization*: A process by a chart or other image is created as a visual representation of the transformed data.

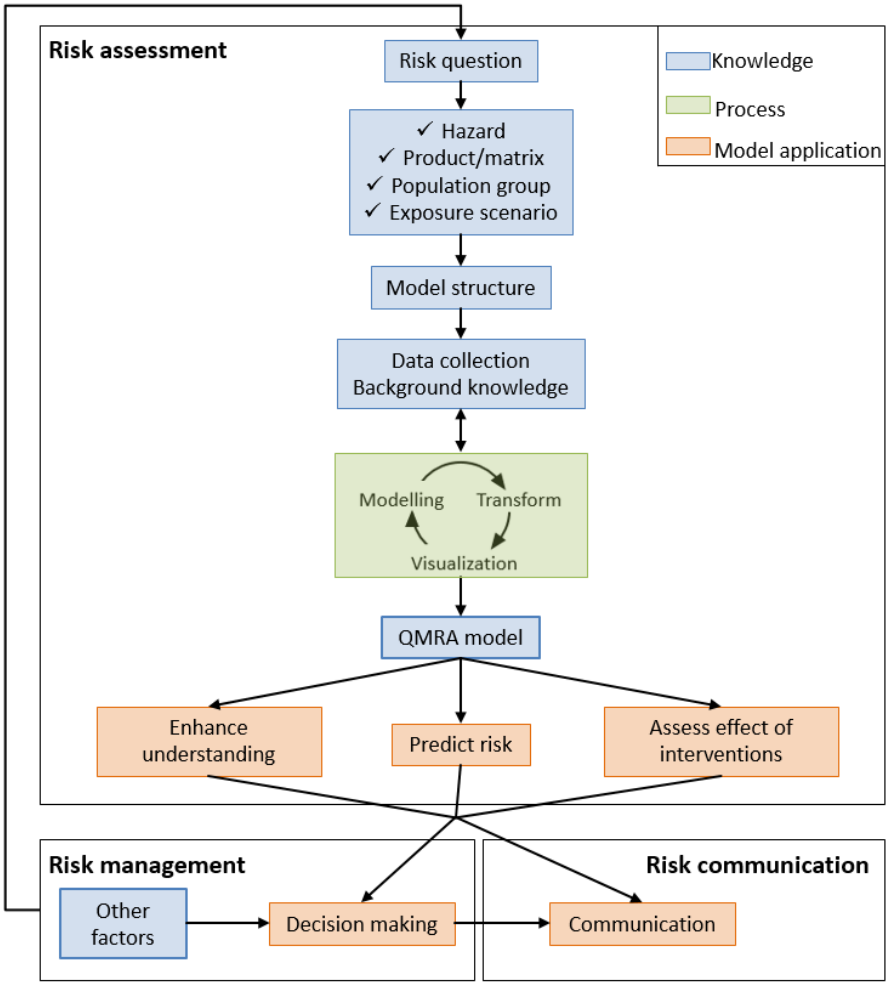


Figure 1. The risk analysis process and the positioning of a quantitative microbial risk assessment (QMRA) modelling process.

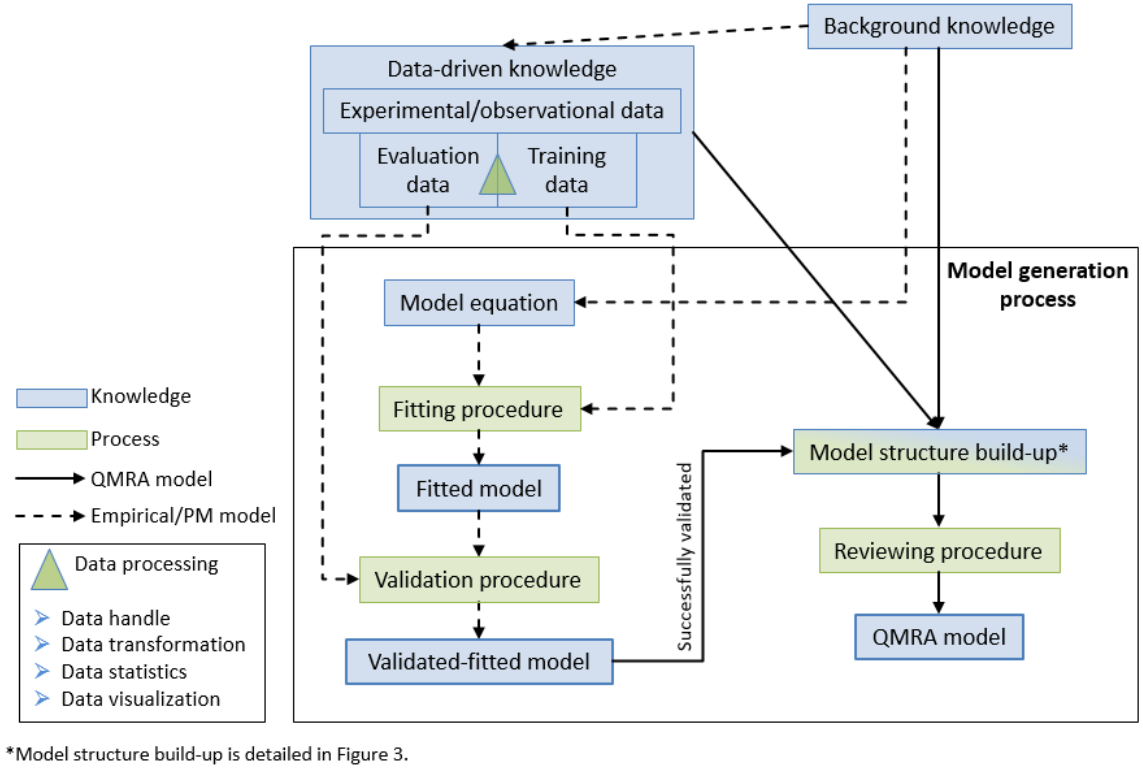


Figure 2. The model generation process for QMRA, Empirical model and PM model.

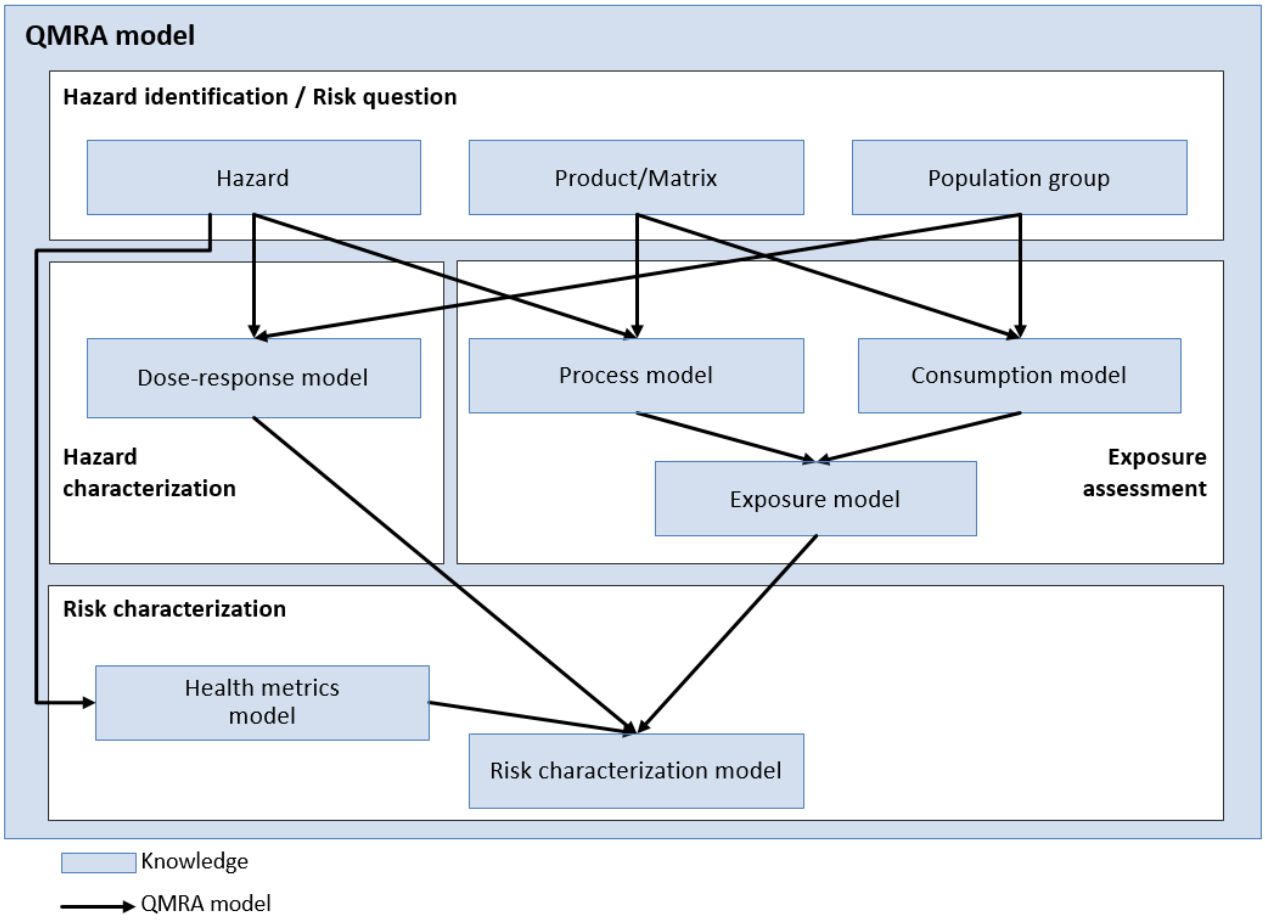


Figure 3. Model structure build-up: QMRA model generation elements outline with frequent model classes.

Table 1. Sources of background knowledge used in model generation process (adapted from Bellet et al. 2012)

Sources	**Examples**
Published scientific studies	1. Contamination data and exposure models 2. Predictive microbial models and data 3. Dose-response data and models.
Unpublished studies	1. Studies and surveys carried out e.g. by the industry, association/society, etc. 2. Laboratory (private or public) databases
National or international official data	1. Food monitoring data 2. Human health surveillance data (e.g. laboratory diagnostic data, historical outbreaks investigations, biomonitoring survey) 3. Food consumption survey or regional diet data 4. Food safety databases.
Surveys and questionnaires	1. Questionnaires sent to manufacturers
Risk assessments (RA)	1. RA carried out by national or international agencies and research institutes
Expert knowledge elicitation and expert opinions	1. When data is not available, estimates can be provided by experts opinion
Events	1. Meetings or conferences

•